

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:14:34 ; Search time 19.84 Seconds  
(without alignments)  
92,401 Million cell updates/sec

Title: US-09-822-110-5

Perfect score: 253  
Sequence: 1 PFIKAKPVSLRLITEAM.....QRELEEDENDEDELDSHT 50Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	253	100.0	491	1	STK3_HUMAN
2	165.5	65.4	487	1	STK4_HUMAN
3	71.5	28.3	312	1	PHO4_YEAST
4	70	27.7	626	1	AL12_ARAHY
5	69	27.3	200	1	HMG1_CHICK
6	69	27.3	201	1	HMG1_CHICK
7	69	27.3	612	1	RPSD_BUCAL
8	68.5	27.1	324	1	SIAL_MOUSE
9	68	26.9	678	1	GARP_PLAUF
10	67.5	26.7	667	1	OS9_HUMAN
11	67.5	26.7	879	1	MCN3_SCHPO
12	67	26.5	178	1	NPM3_HUMAN
13	67	26.5	416	1	CRCR_RAT
14	67	26.5	418	1	YSE2_CABEL
15	66.5	26.3	879	1	SPIO_HUMAN
16	66.5	26.3	890	1	IF2_ECOLI
17	66	26.1	598	1	ABBI_RAT
18	66	26.1	706	1	NUCL_MOUSE
19	66	26.1	1121	1	MYT1_HUMAN
20	65	25.7	182	1	YADA_SCHPO
21	65	25.7	310	1	SIAL_BOVIN
22	65	25.7	976	1	VP41_ARATH
23	64.5	25.5	721	1	YCF2_OENPI
24	64	25.3	180	1	HMG1_CRIGR
25	64	25.3	209	1	HMG2_PIG
26	64	25.3	214	1	HMG1_MOUSE
27	64	25.3	320	1	SIAL_RAT
28	64	25.3	416	1	CRTC_MOUSE
29	64	25.3	639	1	GLCX_SOYBN
30	64	25.3	712	1	NUCL_RAT
31	63.5	25.1	1161	1	NRDC_RAT
32	63	24.9	240	1	SNF_YEAST
33	63	24.9	299	1	NPM_XENLA

34	63	24.9	399	1	MYC_CARAU
35	63	24.9	483	1	CBF5_YEAST
36	63	24.9	630	1	YCF2_OENVI
37	63	24.9	631	1	COB1_PSEDE
38	62.5	24.7	337	1	ATH2_MOUSE
39	62.5	24.7	382	1	NDF2_HUMAN
40	62.5	24.7	417	1	CRTC_HUMAN
41	62.5	24.7	520	1	GAG_GALV
42	62.5	24.7	537	1	GAG_MLYRD
43	62.5	24.7	571	1	YK18_YEAST
44	62.5	24.7	830	1	GYRA_CLOAB
45	62	24.5	209	1	HMG2_MOUSE

## ALIGNMENTS

RESULT	ID	STANDARD	PRT	491 AA
1	STK3_HUMAN			
AC	Q13188; Q15801; Q15445;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DE	SERINE/THREONINE PROTEIN KINASE 3 (EC 2.7.1.37) (STE20-LIKE KINASE			
DE	MST2) (MST-2) (MAMMALIAN STE20-LIKE PROTEIN KINASE 2)			
DE	(SERINE/THREONINE PROTEIN KINASE KRS-1).			
GN	STE20-LIKE KINASE			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96144292; PubMed=8566796;			
RA	Creasy C.L., Chernoff J.,			
RT	"Cloning and characterization of a member of the MST subfamily of			
RT	Ste20-like kinases."			
RT	Gene=167:303-306(1995).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=96413604; PubMed=8816758;			
RA	Taylor L.K., Wang H.C., Erikson R.L.,			
RT	"Newly identified stress-responsive protein kinases, Krs-1 and Krs-			
RT	2."			
RT	Proc. Natl. Acad. Sci. U.S.A. 93:10099-10104(1996)			
RL	[3]			
RP	SEQUENCE OF 96-203 FROM N.A.			
RX	MEDLINE=94100173; PubMed=8274451;			
RA	Schultz S.J., Nigg E.A.,			
RT	"Identification of 21 novel human protein kinases, including 3 members			
RT	of a family related to the cell cycle regulator nima of Aspergillus			
RT	nidulans."			
RL	Cell Growth Differ. 4:821-830(1993).			
CC	- FUNCTION: OXIDANT STRESS-ACTIVATED SERINE/THREONINE KINASE THAT			
CC	MAY PLAY A ROLE IN THE RESPONSE TO ENVIRONMENTAL STRESS (BY			
CC	SIMILARITY).			
CC	- CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.			
CC	- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).			
CC	- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN ADULT KIDNEY,			
CC	SKELETAL AND PLACENTA TISSUES AND AT VERY LOW LEVELS IN ADULT			
CC	HEART, LUNG AND BRAIN TISSUES.			
CC	- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.			
CC	STE20 SUBFAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isdb.ch/announce/">http://www.isdb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isdb.ch">license@isdb.ch</a> ).			
CC				

```
Query Match      100.0%; Score:253; DB 1; Length 491;
Best Local Similarity 100.0%; Pred. No. 4,3e-17;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

RESULT 2  
STK4\_HUMAN

20-AUG-2001 (Rel.: 40, Created)  
20-AUG-2001 (Rel.: 40, Last sequence update)  
20-AUG-2001 (Rel.: 40, Last annotation update)  
SERINE/THREONINE PROTEIN KINASE 4 (EC 2.7.1.37) (STE20-LIKE KINASE)  
MS1) (MS1-1) (MAMMALIAN STE20-LIKE PROTEIN KINASE 1)  
(SERINE/THREONINE PROTEIN KINASE KRS-2).  
SK4 OR MS1.

[1]  
SEQUENCE FROM N.A.  
MEDLINE=95394929; PubMed=7665586;  
Cresay C.L., Chernoff J.,  
"Cloning and characterization of a human  
protein kinase with homology  
to Ste20.",  
J. Biol. Chem. 270:21695-21700(1995).

121 SEQUENCE FROM N.A.  
MEDLINE=96413604; PubMed=8816758;  
TAYLOR L.K., Wang, H.C., Erikson R.L.;  
"Newly identified stress-responsive protein kinases, Krs-1 and Krs  
2.";  
Proc. Natl. Acad. Sci. U.S.A. 93:10099-10104(1996).  
[3] SEQUENCE OF 1-435 FROM N.A.

Submitted (MAR-2000) to the EMBL/GenBank/DBU databases.  
-1- CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.  
-1- ENZYME REGULATION: THE C-TERMINAL NON-CATALYTIC REGION INHIBITS THE KINASE ACTIVITY.

CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -I- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.  
CC -I- PM: AUTOPHOSPHORYLATED ON SERINE AND THREONINE RESIDUES.  
CC -I- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC SEZ20 SUBFAMILY.  
CC  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC

DR	EMBL:	U18297;	AAA83254.1;	-
DR	EMBL:	O60207;	AAB17262.1;	-
DR	EMBL:	AL109839;	CAB89421.1;	-
DR	HSSP:	P24941;	IHCL.	
DR	MIM:	604965;	-	
DR	InterPro:	IPR000719;	Euk_pkinase.	
DR	InterPro:	IPR002290;	Ser_thr_kin_actsite.	
DR	InterPro:	IPR001245;	Tyr_kin.	
DR	Pfam:	PF00069;	pkinaase; 1.	
DR	PRINTS:	PR00109;	TYRKINASE.	
DR	SMART:	SM00220;	S_TKC_1.	
DR	PROSITE:	PS00107;	PROTEIN_KINASE_ATP; 1.	
DR	PROSITE:	PS00108;	PROTEIN_KINASE_ST; FALSE_NEG.	
DR	PROSITE:	PS50011;	PROTEIN_KINASE_DOM; 1.	
DR	Kw	Transferrase; Serine/threonine-protein kinase;	ATP-binding.	
DR	DOMAIN	NP_BIND	30	281
DR	FFT	BINDING	56	44
DR	FFT	ACT_SITE	59	59
DR	FFT	DOMAIN	149	149
DR	FFT	CONFLICT	373	378
DR	FFT	CONFLICT	222	222
DR	FFT	CONFLICT	312	312
DR	SEQUENCE	487 AA;	55630 MW;	150758BC5F77D5C CRC64;

Query Match	Score	DB 1:	Length
Best Local Similarity	65.48;		487;
	63.79;		
	63.79;		

Matches: 32; Conservative: 12; Mismatches: 6; Indels: 1; Gaps: 1.

RESULT	3		
PH04_YEAST			
PH04_YEAST			
D070370			
STANDARD;			
PRT;			
312 AA			

01-APR-1988 (Rel. 07, Created)  
01-FEB-1991 (Rel. 17, last sequence update)  
15-DEC-1998 (Rel. 37, last annotation update)  
PHOSPHATE SYSTEM POSITIVE REGULATORY PROTEIN PHO4,  
PHO4 OR YFR034C.  
*Saccharomyces cerevisiae* (Baker's yeast).

SEQUENCE FROM N.A.  
MEDLINE=66176785; PubMed=3008105;  
Legrain M., de Wilde M., Hilger F. :  
"Isolation, physical characterization and expression analysis of the  
Saccharomyces cerevisiae positive regulatory gene PHO4." ;  
Nucleic Acids Res. 14:3059-3073(1986).  
[1]  
NCBI\_TaxID=4932;  
Saccharomycetes;  
Saccharomycetales;  
Saccharomycetaceae;  
Saccharomyces;  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
C Saccharomycetales;  
X Saccharomycetaceae;  
X Saccharomyces;  
P  
N

[2] SEQUENCE FROM N.A. MEDLINE-89364688; PubMed=2605053; Yoshida K., Kuromitsu Z., Ogawa N., Oshima Y.; "Mode of expression of the positive regulatory genes *pho2* and *pho4* of *S. cerevisiae* in response to phosphate starvation"; *Yeast* 1990; 6: 111-119.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:04:59 ; Search time 26.93 Seconds

(without alignments)  
41.781 Million cell updates/sec

Title: US-09-822-110-5

Perfect score: 253

Sequence: 1 PFKNKAPVSLRDILITEAM.....QRELEEEENSDELDLSHT 50

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfill.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	166.5	65.8	487	2	US-08-712-709-8
2	166.5	65.8	487	3	US-09-111-444-8
3	166.5	65.8	487	4	US-09-541-228-8
4	67	26.5	416	2	US-09-211-930-11
5	67	26.5	416	3	US-09-340-993-11
6	67	26.5	416	4	US-09-468-442-11
7	67	26.5	1162	2	US-08-728-333A-2
8	65	25.7	416	2	US-09-211-930-3
9	65	25.7	416	3	US-09-340-993-3
10	65	25.7	416	4	US-09-468-442-3
11	65	25.7	759	1	US-08-676-967-1
12	65	25.7	759	1	US-08-676-967-1
13	65	25.7	759	2	US-09-098-487-1
14	64	25.3	1719	2	US-08-459-568-4
15	64	25.3	1719	2	US-08-399-411-4
16	64	25.3	1719	3	US-08-516-859A-4
17	62.5	24.7	379	1	US-08-552-142A-11
18	62.5	24.7	381	1	US-08-910-973-11
19	62.5	24.7	382	3	US-09-224-332-9
20	62	24.5	353	3	US-08-688-988-31
21	62	24.5	781	2	US-08-373-134D-2
22	62	24.5	781	2	US-09-114-637-2
23	62	24.5	3135	1	US-08-323-170B-2
24	61	24.1	270	2	US-08-853-743-5
25	61	24.1	270	3	US-09-185-370-5
26	61	24.1	285	4	US-09-282-305-18
27	61	24.1	905	2	US-08-574-959A-9

28	61	24.1	905	4	US-09-357-014-9	Sequence 9, Appli
29	61	24.1	1135	2	US-08-574-959A-7	Sequence 7, Appli
30	61	24.1	1135	4	US-09-357-014-7	Sequence 9, Appli
31	60.5	23.9	675	1	US-08-317-522A-9	Sequence 9, Appli
32	60.5	23.9	675	1	US-08-439-818A-9	Sequence 9, Appli
33	60.5	23.9	675	2	US-08-751-965-9	Sequence 9, Appli
34	60.5	23.9	675	2	US-08-738-975-9	Sequence 9, Appli
35	60.5	23.9	675	2	US-08-728-626-9	Sequence 9, Appli
36	60.5	23.9	675	3	US-08-808-599A-9	Sequence 9, Appli
37	60	23.7	240	3	US-08-114-555A-8	Sequence 8, Appli
38	60	23.7	240	3	US-08-559-397A-14	Sequence 14, Appli
39	60	23.7	303	4	US-09-045-973-1	Sequence 1, Appli
40	60	23.7	522	4	US-09-075-272-3	Sequence 3, Appli
41	60	23.7	622	4	US-09-075-272-2	Sequence 2, Appli
42	60	23.7	887	1	US-07-867-106-3	Sequence 3, Appli
43	60	23.7	1085	1	US-08-431-080-28	Sequence 28, Appli
44	60	23.7	1085	2	US-08-938-534-28	Sequence 28, Appli
45	59.5	23.5	310	2	US-08-943-600A-3	Sequence 3, Appli

#### ALIGNMENTS

RESULT 1  
US-08-712-709-8  
Sequence 8, Application US/08712709  
Patent No. 5863780  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Guegler, Karl J.  
APPLICANT: Hawkins, Phillip R.  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/712,709  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0118 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 487 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1117791  
US-08-712-709-8

Query Match 65.8% Score 166.5; DB 2; Length 487;  
Best local similarity 62.7%; Pred. No. 2.5e12;  
Matches 32; Conservative 12; Mismatches 6; Indels 1; Gaps 1;  
Cy 1 PFKNKAPVSLRDILITEAMIKAKRHDEQREL-EEEEENSDELDLSHT 50



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:06:00 ; Search time 54.15 Seconds

(without alignments)  
68.396 Million cell updates/sec

Title: US-09-822-110-5

Perfect score: 253  
Sequence: 1 PPIKNAKPVSIILRLITLTEM.....ORELEEEENSDDELDSHT 50

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSB8/gcgdata/geneseq/AA1980.DAT:\*  
2: /SIDSB8/gcgdata/geneseq/AA1981.DAT:\*  
3: /SIDSB8/gcgdata/geneseq/AA1982.DAT:\*  
4: /SIDSB8/gcgdata/geneseq/AA1983.DAT:\*  
5: /SIDSB8/gcgdata/geneseq/AA1984.DAT:\*  
6: /SIDSB8/gcgdata/geneseq/AA1985.DAT:\*  
7: /SIDSB8/gcgdata/geneseq/AA1986.DAT:\*  
8: /SIDSB8/gcgdata/geneseq/AA1987.DAT:\*  
9: /SIDSB8/gcgdata/geneseq/AA1988.DAT:\*  
10: /SIDSB8/gcgdata/geneseq/AA1989.DAT:\*  
11: /SIDSB8/gcgdata/geneseq/AA1990.DAT:\*  
12: /SIDSB8/gcgdata/geneseq/AA1991.DAT:\*  
13: /SIDSB8/gcgdata/geneseq/AA1992.DAT:\*  
14: /SIDSB8/gcgdata/geneseq/AA1993.DAT:\*  
15: /SIDSB8/gcgdata/geneseq/AA1994.DAT:\*  
16: /SIDSB8/gcgdata/geneseq/AA1995.DAT:\*  
17: /SIDSB8/gcgdata/geneseq/AA1996.DAT:\*  
18: /SIDSB8/gcgdata/geneseq/AA1997.DAT:\*  
19: /SIDSB8/gcgdata/geneseq/AA1998.DAT:\*  
20: /SIDSB8/gcgdata/geneseq/AA2000.DAT:\*  
21: /SIDSB8/gcgdata/geneseq/AA2001.DAT:\*  
22: /SIDSB8/gcgdata/geneseq/AA2001.DAT:\*

-Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	167.5	66.2	487	20	AAV21674 Human Ste20 homolo
2	74	29.2	524	18	AAW32096 Miniature swine re
3	74	29.2	524	22	AAW3285 Retroviral protein
4	71	28.1	142	21	AAW41902 Arabidopsis thalia
5	71	28.1	261	21	AAW41901 Arabidopsis thalia
6	71	28.1	280	21	AAW41900 Arabidopsis thalia
7	70	27.7	415	20	AAV40913 Ara h 1 allergen p
8	70	27.7	626	18	AAW21550 Peanut allergen Ar
9	70	27.7	626	20	AAV15244 Peanut allergen, A
10	70	27.7	626	20	AAV25657 Peanut allergen 11
11	70	27.7	626	22	AAW04706 Anaphylactic anti

12	70	27.7	634	21	AAW33599 Modified Ara h 1 a
13	70	27.7	634	22	AAW04709 Modified anaphylac
14	70	27.7	634	22	AAW05034 Modified anaphylac
15	69	27.3	117	22	AAW36518 Peptide #10555 enc
16	69	27.3	524	18	AAW39271 Porcine retrovirus
17	69	27.3	524	18	AAW32091 Porcine retrovirus
18	69	27.3	524	22	AAW70629 Porcine endogenous
19	69	27.3	524	22	AAW73282 Defective retrovir
20	67	26.5	162	21	AAW29982 Arabidopsis thalia
21	67	26.5	214	21	AAW21047 Human nucleic acid
22	67	26.5	241	21	AAW29981 Arabidopsis thalia
23	67	26.5	247	21	AAW29980 Arabidopsis thalia
24	67	26.5	416	20	AAW22651 A murine signal tr
25	67	26.5	416	21	AAW21673 Murine 5e.new poly
26	67	26.5	416	21	AAW82277 Mouse protein poly
27	67	26.5	1162	21	AAV6255 Kaposi's sarcoma-a
28	67	26.5	1162	21	AAV58500 HHV8 ORF 73 protel
29	67	26.5	1162	22	AAW62331 Amino acid sequenc
30	66.5	26.3	993	22	AAW59437 Human polypeptide
31	66.5	26.3	1312	21	AAW76884 Retinoblastoma bln
32	66.5	26.3	1312	22	AAW38967 Human polypeptide
33	66	26.1	211	21	AAW3517 Arabidopsis thalia
34	66	26.1	219	21	AAW33516 Arabidopsis thalia
35	66	26.1	309	21	AAW23515 Arabidopsis thalia
36	66	26.1	562	16	AAW70491 Leucocytoczoan prot
37	66	26.1	1151	20	AAW5039 Human N-arginine d
38	65.5	25.9	1284	21	AAW37724 Arabidopsis thalia
39	65.5	25.9	329	21	AAW37723 Arabidopsis thalia
40	65.5	25.9	360	21	AAW37722 Arabidopsis thalia
41	65	25.7	65	22	AAW15961 Peptide #2395 enco
42	65	25.7	65	22	AAW28463 Peptide #2300 enco
43	65	25.7	65	22	AAW03698 Peptide #2380 enco
44	65	25.7	228	22	AAW63361 Human breast cance
45	65	25.7	231	22	AAW94376 Human protein sequ

#### ALIGNMENTS

RESULT 1	AAV21674	AAV21674 standard; Protein: 487 AA.
ID	AAV21674	
AC	AAV21674	
XX		
DT	18-AUG-1999 (first entry)	
XX		
DE	Human Ste20 homologue polypeptide S201.	
XX		
KW	Sterile 20; Ste20; serine/threonine kinase; C12.2bs; 5e.new; murine; erythroleukemia; mitogenic signalling; cell differentiation; metastasis; signaling pathway; cancer; ischemic stroke; heart disease; inflammation.	
KW		
OS	Homo sapiens.	
XX		
PN	W09929857-A1.	
XX		
PD	17-JUN-1999.	
XX		
PF	09-DEC-1998; 98W0-US26116.	
XX		
PR	09-DEC-1997; 97US-0069078.	
XX		
PA	(CHIL-) CHILDRENS MEDICAL CENT.	
XX		
PI	Agarwal S, Best J; Vail B, Zon LI;	
XX		
DR	WPI: 1999-385605/32.	
XX		
PT	Murine homologues of the Sterile 20 family of serine/threonine kinases, useful in elucidation of mitogenic signalling pathways	
XX		
PS	Disclosure: Fig 4; 37p; English.	

xx The invention provides two members of the Sterile 20 (Ste20) family of  
 CC serine/threonine kinases (C12.2bs and 5e.new). Host cells transfected  
 CC with vectors comprising the nucleic acids encoding the polypeptides are  
 CC used for their recombinant expression. The Ste20 kinases, isolated from a  
 CC murine erythroleukemia cDNA library, are regulatory molecules involved in  
 CC mitogenic signaling as well as other cellular phenomena such as  
 CC morphology and motility. These activities are important factors in  
 CC development, cell differentiation, cancer and metastases. Manipulation of  
 CC the C12.2bs and 5e.new proteins or polynucleotides are useful for  
 CC manipulation of the signaling pathways involved and will allow  
 CC development of reagents to modulate these signaling pathways. The kinases  
 CC are also important for ischemic stroke, heart disease and inflammation.  
 CC The antibodies can be used to inhibit the activity of the kinase,  
 CC especially in vitro and in cell extracts, as well as identify the kinases  
 CC in immunoassays. The present sequence represents a S201 polypeptide, a  
 CC human homologue of Ste20.

xx Sequence 487 AA:

Query Match 66.2% Score 167.5; DB 20; Length 487;  
 Best Local Similarity 64.7%; Pred. No. 5.4e-12;  
 Matches 33; Conservative 11; Mismatches 6; Indels 1; Gaps 1;

OY 1 PFIKNAKPVSLDLITEAMEIKAKRHDEQOREL-EEEEENSDDEDLSHT 50  
 DB 279 pfkakskysilrldineamdvlkkrqeqvqdqddenseedmdsgt 329

## RESULT 2

AAW32096 AAW32096 standard; Protein; 524 AA.

AC AAW32096;

DT 09-FEB-1998 (first entry)

DE Miniature swine retrovirus GAG;protein.

XX Retrovirus; porcine; GAG protein; xenotransplantation; infectious;

KW provirus; organ transplant; donor; activated virus; PCR.

OS Porcine retrovirus.

XX Key Location/Qualifiers

FT Protein 1..524

XX /Label= GAG\_protein

PN WO9721836-A1.

PD 19-JUN-1997.

XX 13-DEC-1996; 96WO-US19680.

PR 14-DEC-1995; 95US-0572645.

PA (GEHO ) GEN HOSPITAL CORP.

PI Fishman JA;

DR WPI; 1997-332804/30.

DR N-PSDB; AAT74884.

XX New nucleic acid from porcine retroviruses - used for detecting

PT viruses in transplant or other tissue and for assessing risk of

XX transmitting infection to graft recipient.

PS Claim 22; Fig 3; 128pp; English.

CC This is a porcine retrovirus from miniature swine containing the coding

CC region for a putative viral GAG protein. This sequence and PCR fragments

CC generated from the sequence (see AAT74812-T74882) could be used to

CC screen organs for porcine retroviruses prior to xenotransplantation.  
 CC Transplantation can increase the likelihood of retroviral activation if  
 CC intact and infectious proviruses are present. The porcine retroviral  
 CC sequence can be used to generate probes to determine the level (e.g.  
 CC copy number) of intact (i.e. potentially replicating) porcine provirus  
 CC sequences in a strain of xenograft transplantation donors. It can be  
 CC used to detect mutations, genetic lesions or viral recombinants and  
 CC also to determine the histological localisation of activated retrovirus.  
 CC using Polymerase Chain Reaction DNA Quantitation (PDQ) on blood  
 CC mononuclear cells, infectivity titration and susceptibility testing can  
 CC be performed. Ultimately animal donors without intact porcine retroviral  
 CC sequences or a lower copy number of viral elements could be selected.

xx Sequence 524 AA:

Query Match 29.2% Score 74; DB 18; Length 524;  
 Best Local Similarity 41.2%; Pred. No. 0.78;  
 Matches 14; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

OY 12 LRDITEAMEIKAKRHDEQORELEEEENSDDE 45  
 DB 406 lrdlykaekykykreteereqrkerereee 439

## RESULT 3

AAB73285 AAB73285 standard; Protein; 524 AA.

AC AAB73285;

DT 23-MAY-2001 (first entry)

DE Retroviral protein #1 found in miniature swine.

XX Retrovirus; graft transplantation; xenotransplantation; miniature swine.

XX Unidentified.

OS US6190861-B1.

PN 20-FEB-2001.

PD 13-DEC-1996; 96US-0766528.

XX 14-DEC-1995; 95US-0572645.

PR (GEHO ) GEN HOSPITAL CORP.

PA Fishman JA;

PI WPI; 2001-256211/26.

DR N-PSDB; AAF77727.

XX Assessing risk of endogenous retroviruses in clinical practice and in

PT xenotransplantation, comprises using probe sequences derived from swine

XX or miniature swine retroviral genome -

PS Disclosure; Fig 3; 127pp; English.

XX The present invention relates to a method for screening a cell or tissue

CC for the presence or expression of a retrovirus (RV), comprising

CC contacting a target nucleic acid from the cell or tissue with a second

CC nucleic acid from the present invention (e.g. AAF77727 or a fragment

CC thereof). The method is useful for RV detection and to assess graft

CC transplantation risk. Screening of animals allows the elimination of

CC donors with active replication of known viruses. Inactive proviruses can

CC be detected and inactivated, allowing identification and elimination of

CC potential human pathogens derived from swine in a manner not possible in

CC the outbred human organ donor population and is important to the

CC development of human xenotransplantation.

xx Sequence 524 AA;

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:14:08 ; Search time 53.55 Seconds

(without alignments)  
136.575 Million cell updates/sec

Title: US-09-822-110-5

Perfect score: 253  
Sequence: 1 PFKNKAPVSIILRLITRAM.....ORELEEEENDEDELDSHT 50

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPTRMBL\_17:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhcc:\*  
8: sp\_mammal:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	244	96.4	321	11	09CW82 mus musculu
2	244	96.4	445	11	060877 mus musculu
3	244	96.4	497	11	09J110 mus musculu
4	238	96.1	491	11	054748 mus musculu
5	167.5	66.2	487	11	09J111 mus musculu
6	165.5	65.4	210	4	09BR32 mus musculu
7	90	35.6	478	5	062571 mus musculu
8	76.5	30.2	323	5	09W117 mus musculu
9	76	30.0	642	5	09VEN3 mus musculu
10	75	29.6	950	13	09YHC9 mus musculu
11	74	29.2	524	6	062704 mus musculu
12	74	29.2	524	6	062706 mus musculu
13	73	28.9	309	5	09J318 mus musculu
14	73	28.9	775	11	09DC65 mus musculu
15	73	28.9	775	11	09DC65 mus musculu
16	72	28.5	1087	5	096923 mus musculu
17	72	28.5	1110	13	091255 mus musculu
18	71.5	28.3	312	3	006859 mus musculu
19	71.5	28.3	312	3	006860 mus musculu

20	71.5	28.3	461	5	061125 dictyosteli
21	71	28.1	292	10	09LVH1 caenorhabdi
22	71	28.1	1309	5	017582 caenorhabdi
23	70.5	27.9	3781	5	09VIS5 drosophila
24	70	27.7	775	5	09W261 drosophila
25	70	27.7	916	12	09CW70 avian adeno
26	69.5	27.5	1100	5	09N330 caenorhabdi
27	69	27.3	524	6	062708 mus musculu
28	69	27.3	664	6	09N117 mus musculu
29	69	27.3	707	11	099K50 mus musculu
30	69	27.3	1520	4	015087 mus musculu
31	69	27.3	1781	4	09UKX0 mus musculu
32	69	27.3	1890	4	09UKW3 mus musculu
33	69	27.3	2073	4	09UKW2 mus musculu
34	69	27.3	2376	12	09Q1X5 mus musculu
35	69	27.3	2376	12	09Q1X3 mus musculu
36	69	27.3	2378	12	09Q1X4 mus musculu
37	68.5	27.1	313	10	09C6H3 mus musculu
38	68.5	27.1	539	5	09VPM5 mus musculu
39	67.5	26.7	376	4	09BR60 mus musculu
40	67.5	26.7	612	4	09BW99 mus musculu
41	67.5	26.7	1171	3	09P3E2 mus musculu
42	67	26.5	1178	4	09UNY6 mus musculu
43	67	26.5	299	10	048722 mus musculu
44	67	26.5	403	5	076961 mus musculu
45	67	26.5	403	5	09XYF8 mus musculu

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	321 AA.
09CW82	09CW82	09CW82		
AC	01-JUN-2001 (TREMBL)	17, Created		
DT	01-JUN-2001 (TREMBL)	17, Last sequence update		
DT	01-JUN-2001 (TREMBL)	17, Last annotation update		
DE	0610042106RIK	PROTEIN (FRAGMENT)		
GN	0610042106RIK			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=KIDNEY;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Kono H., Kasukawa T., Saito R.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada S.,			
RA	Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bolt C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Guinacich S., Hillid, Hofmann M., Hume D.A., Kamuya M., Lee N.H.,			
RA	Lions P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,			
RA	Nordone P., Ring B., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,			
RA	Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa T., Kawaji H., Kotsuki S.,			
RA	Hayashizaki Y.,			
RT	"Functional annotation of a full-length mouse cDNA collection";			
RL	Nature 409:685-690(2001).			
DR	EMBL: AK002914; BAB22453.1;			
DR	MGD: MGI:1914102; 0610042106RIK.			
DR	InterPro: IPR000719; Euk_pkinase.			
DR	InterPro: IPR002290; Ser_thr_kin_actsite.			
DR	Pfam: PF00069; pkinase; I.			

DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Transferase.  
FT NON\_TER 1 1  
SQ SEQUENCE 321 AA; 37201 MW; 2696FD4E919F53A4 CRC64;

Query Match  
Best Local Similarity 96.4%; Score 244; DB 11; Length 321;  
Matches 48; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PF1KNAKPVSIIRDLITTEAMEIKAKRHDEQORELEEEENSDELDLSHT 50  
DB 100 PF1KNAKPVSIIRDLITTEAMEIKAKRHDEQORELEEEENSDELDLSHT 149

RESULT 2  
O60877 PRELIMINARY; PRT; 445 AA.

O60877:  
01-NOV-1996 (TREMBLrel. 01, Created)  
01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PROTEIN KINASE MESS1.  
GN STK3 OR MESS1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LIVER;  
RA Han J.;  
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.  
DR EMBL; U28726; AA075300.1;  
DR HSSP; P00518; IPRK.  
DR MGD; MGI:1928487; SCK3.  
DR InterPro; IPR000719; Euk\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
DR Pfam; PF00069; pkinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Transferase.  
SQ SEQUENCE 445 AA; 50407 MW; 78936B28DA616F05 CRC64;

Query Match  
Best Local Similarity 96.4%; Score 244; DB 11; Length 445;  
Matches 48; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PF1KNAKPVSIIRDLITTEAMEIKAKRHDEQORELEEEENSDELDLSHT 50  
DB 276 PF1KNAKPVSIIRDLITTEAMEIKAKRHDEQORELEEEENSDELDLSHT 325

RESULT 3  
O9J110 PRELIMINARY; PRT; 497 AA.

O9J110:  
01-OCT-2000 (TREMBLrel. 15, Created)  
01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE STE20-LIKE KINASE MST2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BAUB/C;  
RA Lee K.-K.; Ohyama T.; Yajima N.; Yonehara S.;  
FT "MST", a physiological caspase substrate, highly sensitizes apoptosis

FT both upstream and downstream of caspase activation."  
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AF271361; AA075790.1;  
DR InterPro; IPR000719; Euk\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
DR InterPro; IPR001245; Tyr\_kin.  
DR Pfam; PF00069; pkinase; 1.  
DR PRINTS; PR00109; TYRKINASE.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; kinase; transferase.  
SQ SEQUENCE 497 AA; 56855 MW; 9CDD365437581665 CRC64;

Query Match  
Best Local Similarity 96.4%; Score 244; DB 11; Length 497;  
Matches 48; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 PF1KNAKPVSIIRDLITTEAMEIKAKRHDEQORELEEEENSDELDLSHT 50  
DB 276 PF1KNAKPVSIIRDLITTEAMEIKAKRHDEQORELEEEENSDELDLSHT 325

RESULT 4  
O54748 PRELIMINARY; PRT; 491 AA.

O54748:  
01-JUN-1998 (TREMBLrel. 06, Created)  
01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE MST2 KINASE.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Aurisicchio L.; Dilauro R.; Zannini M.;  
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.  
DR EMBL; A1001529; CAA04814.1;  
DR HSSP; P00518; IPRK.  
DR InterPro; IPR000719; Euk\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kin\_actsite.  
DR Pfam; PF00069; pkinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Transferase.  
SQ SEQUENCE 491 AA; 56121 MW; 90FA6B020E7FFD62 CRC64;

Query Match  
Best Local Similarity 94.1%; Score 238; DB 11; Length 491;  
Matches 46; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 PF1KNAKPVSIIRDLITTEAMEIKAKRHDEQORELEEEENSDELDLSHT 50  
DB 276 PF1KNAKPVSIIRDLITTEAMEIKAKRHDEQORELEEEENSDELDLSHT 325

RESULT 5  
O9J111 PRELIMINARY; PRT; 487 AA.

O9J111:  
01-OCT-2000 (TREMBLrel. 15, Created)  
01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE STE20-LIKE KINASE MST1.  
GN STK4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.



18 EAMEIKAKRHDEQORELEEESEENSDEDELDST 30



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:14:32; Search time 19.84 Seconds

(without alignments)  
29.568 Million cell updates/sec

Title: US-09-822-110-3

Sequence: 1 EIKAKRHDEQRELEE 16

Scoring table: BIOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81	100.0	491	1	STK3_HUMAN
2	46	56.8	487	1	STK4_HUMAN
3	45	55.6	329	1	MLO2_SCHPO
4	45	55.6	1427	1	REST_HUMAN
5	44	54.3	1537	1	MTDM_CHICK
6	42	51.9	197	1	DEMI_PHYSA
7	42	51.9	198	1	DEMI_PHYSA
8	42	51.9	482	1	UZR2_HUMAN
9	42	51.9	890	1	IF2_ECOLI
10	42	51.9	896	1	IF2_ECOLI
11	42	51.9	897	1	IF2_ECOLI
12	42	51.9	1240	1	YF10_YEAST
13	41.5	51.2	458	1	YF10_YEAST
14	41.5	51.2	536	1	DNK1_THEMA
15	41	50.6	143	1	Y082_RICPR
16	41	50.6	180	1	GLUC_BOVIN
17	41	50.6	180	1	GLUC_BOVIN
18	41	50.6	180	1	GLUC_BOVIN
19	41	50.6	180	1	GLUC_BOVIN
20	41	50.6	180	1	GLUC_BOVIN
21	41	50.6	180	1	GLUC_BOVIN
22	41	50.6	180	1	GLUC_BOVIN
23	41	50.6	180	1	GLUC_BOVIN
24	41	50.6	180	1	GLUC_BOVIN
25	40	49.4	388	1	PAM_YEAST
26	40	49.4	388	1	PAM_YEAST
27	40	49.4	866	1	FIBA_HUMAN
28	40	49.4	1150	1	YK01_CAEEL
29	40	49.4	2128	1	SPCB_MOUSE
30	39.5	48.8	385	1	CNS1_YEAST
31	39	48.1	338	1	STRA_AOAE
32	39	48.1	389	1	M49_STRPY
33	39	48.1	404	1	VE2_HPV60

34	39	48.1	444	1	STD1_YEAST
35	39	48.1	849	1	PRO5_YEAST
36	39	48.1	892	1	IF2_SALTY
37	39	48.1	945	1	PNP2_YEAST
38	39	48.1	1010	1	WNT5_DROME
39	39	48.1	1052	1	BU1B_MOUSE
40	39	48.1	1394	1	E75B_DROME
41	39	48.1	2230	1	G0G4_HUMAN
42	38.5	47.5	427	1	YF09_MYCPN
43	38.5	47.5	539	1	PNUT_DROME
44	38.5	47.5	614	1	RA26_SCHPO
45	38	46.9	336	1	RPOA_THEMA

## ALIGNMENTS

RESULT 1

ID	STK3_HUMAN	STANDARD	PRT	491 AA
AC	Q13188; Q15801; Q15445;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	SERINE/THREONINE PROTEIN KINASE 3 (EC 2.7.1.37) (STE20-LIKE KINASE MST2) (MST-2) (MAMMALIAN STE20-LIKE PROTEIN KINASE 2)			
DE	(SERINE/THREONINE PROTEIN KINASE KRS-1)			
GN	STK3 OR MST2			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBI:Taxid:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-96144292; PubMed-8566796;			
RA	Creasy C.L., Chernoff J.,			
RT	"Cloning and characterization of a member of the MST subfamily of Ste20-like kinases."			
RT	Gene 167:303-306(1995).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-96413604; PubMed-8816758;			
RA	Taylor L.K., Wang H.C., Erikson R.L.,			
RT	"Newly identified stress-responsive protein kinases, KRS-1 and KRS-2."			
RT	"Identification of 21 novel human protein kinases, including 3 members of a family related to the cell cycle regulator nimA of Aspergillus nidulans."			
RT	Cell Growth Differ. 4:821-830(1993).			
CC	FUNCTION: OXIDANT STRESS-ACTIVATED SERINE/THREONINE KINASE THAT MAY PLAY A ROLE IN THE RESPONSE TO ENVIRONMENTAL STRESS (BY SIMILARITY).			
CC	CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.			
CC	SUBCELLULAR LOCATION: CYTOSOL; PLASMA MEMBRANE (BY SIMILARITY).			
CC	TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN ADULT KIDNEY, SKELETAL AND PLACENTA TISSUES AND AT VERY LOW LEVELS IN ADULT HEART, LUNG AND BRAIN TISSUES.			
CC	SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. STE20 SUBFAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			

DR EMBL: U26424; AAC50386.1; -  
 DR EMBL: U60206; AAB17262.1; -  
 DR EMBL: Z25422; CAAB0909.1; -  
 DR HSSP: P00518; 1PHK.  
 DR MIM: 605030; -  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
 DR Pfam: PF00069; Pkinase.1.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR SMART: SM00220; S\_TKC; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP\_1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST\_FALSE\_NEG.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM\_1.  
 DR Trasnferase: Serine/threonine-protein kinase; ATP-binding.  
 DR DOMAIN 27 278 PROTEIN KINASE.  
 FT NP\_BIND 33 41 ATP (BY SIMILARITY).  
 FT BINDING 56 56 ATP (BY SIMILARITY).  
 FT ACT\_SITE 146 146 BY SIMILARITY.  
 FT DOMAIN 308 314 POLY-GLU.  
 FT CONFLICT 96 98 MIV -> YLY (IN REF. 3).  
 FT CONFLICT 121 121 D -> Y (IN REF. 3).  
 FT CONFLICT 203 203 D -> G (IN REF. 3).  
 FT CONFLICT 303 303 D -> E (IN REF. 2).  
 FT CONFLICT 332 334 GEC -> ESV (IN REF. 2).  
 SQ SEQUENCE 491 AA; 56261 MW; 9CA3B0644F3C14A9 CRC64;

Query Match 100.0%; Score 81; DB 1; Length 491;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-05;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 EIKAKRHDEOORELEE 16  
 |||||  
 Db 296 EIKAKRHDEOORELEE 311

RESULT 2  
 ID STRK4\_HUMAN STANDARD; PRT; 487 AA.  
 AC Q13043; Q15802; Q9NTZ4;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SERINE/THREONINE PROTEIN KINASE 4 (EC 2.7.1.37) (STE20-LIKE KINASE  
 DE MST1) (MST-1) (MAMMALIAN STE20-LIKE PROTEIN KINASE 1)  
 DE (SERINE/THREONINE PROTEIN KINASE KRS-2).  
 DE STRK4 OR MST1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95394929; Pubmed=7665586;  
 RA Creasy C.L., Chernoff J.;  
 RT "Cloning and characterization of a human protein kinase with homology  
 RT to Ste20."  
 RL J. Biol. Chem. 270:21695-21700(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96413604; Pubmed=8816758;  
 RA Taylor L.K., Wang H.C., Erikson R.L.;  
 RT "Newly identified stress-responsive protein kinases, Krs-1 and Krs-  
 RT 2."  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:10099-10104(1996).  
 RN [3]  
 RP SEQUENCE OF 1-435 FROM N.A.  
 RA Laird G.;  
 CC CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.  
 CC ENZYME REGULATION: THE C-TERMINAL NON-CATALYTIC REGION INHIBITS  
 CC THE KINASE ACTIVITY.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: UNOUTGOSSLY EXPRESSED.  
 CC -1- PTM: AUTOPHOSPHORYLATED ON SERINE AND THREONINE RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC STE20 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: U18297; AAB83254.1; -  
 DR EMBL: U60207; AAB17262.1; -  
 DR EMBL: AL109839; CAB89421.1; -  
 DR HSSP: P24941; 1HCL.  
 DR MIM: 604965; -  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
 DR Pfam: PF00069; Pkinase.1.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR SMART: SM00220; S\_TKC; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP\_1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST\_FALSE\_NEG.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM\_1.  
 DR Trasnferase: Serine/threonine-protein kinase; ATP-binding.  
 DR DOMAIN 30 281 PROTEIN KINASE.  
 FT NP\_BIND 36 44 ATP (BY SIMILARITY).  
 FT BINDING 59 59 ATP (BY SIMILARITY).  
 FT ACT\_SITE 149 149 BY SIMILARITY.  
 FT DOMAIN 373 378 POLY-GLU.  
 FT CONFLICT 222 222 P -> R (IN REF. 1).  
 FT CONFLICT 312 312 V -> M (IN REF. 1).  
 SQ SEQUENCE 487 AA; 55630 MW; 150758EBC5F77D5C CRC64;

Query Match 56.88; Score 46; DB 1; Length 487;  
 Best Local Similarity 43.88; Pred. No. 10;  
 Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 EIKAKRHDEOORELEE 16  
 |||||  
 Db 299 DVKLKROESQOREVDQ 314

RESULT 3  
 ID MLO2\_SCHPO STANDARD; PRT; 329 AA.  
 AC Q09329;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN MLO2.  
 DE MLO2 OR SPBC4.05.  
 GN Schizosaccharomyces pombe (fission yeast).  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=97128260; Pubmed=8972853;  
 RA Javerzat J.-P., Cranston G., Allshire R.A.;  
 RT "Fission yeast genes which disrupt mitotic chromosome segregation  
 RT when overexpressed."  
 RL Nucleic Acids Res. 24:4676-4683(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:14:07 ; Search time 53.55 Seconds  
(without alignments)  
43.704 Million cell updates/sec

Title: us-09-822-110-3

Perfect score: 81

Sequence: 1 EIKAKRHDEQRELEE 16

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues  
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SPREMBL\_17:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp-organella:\*  
10: sp-phage:\*  
11: sp-plant:\*  
12: sp-rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	77	95.1	321	11	Q9CWB82
2	77	95.1	445	11	Q60877
3	77	95.1	497	11	Q9J110
4	74	91.4	491	11	Q54748
5	48	59.3	799	4	Q60302
6	47	58.0	2357	5	Q901M8
7	46.5	57.4	222	2	Q9R1D9
8	46.5	57.4	233	2	Q68746
9	46.5	57.4	210	4	Q9BR32
10	45	55.6	487	11	Q9J111
11	45	55.6	565	5	Q9VWD4
12	45	55.6	625	10	Q9SPJ3
13	45	55.6	666	10	Q9SPL5
14	45	55.6	666	10	Q9SPL4
15	45	55.6	1320	11	Q9J125
16	44.5	54.9	3328	5	Q9VUB5
17	44.5	54.9	455	3	Q12080
18	44	54.3	437	11	Q9C041
19	44	54.3	1621	5	Q01900

20	43	53.1	344	5	Q76741	Q76741 dictyosteli
21	43	53.1	363	4	Q9NV16	Q9NV16 homo sapien
22	43	53.1	444	2	Q9RV42	Q9RV42 deinococcus
23	43	53.1	477	5	Q9NAJ9	Q9NAJ9 caenorhabdi
24	43	53.1	499	10	Q23211	Q23211 arabidopsis
25	43	53.1	522	10	Q9FXB5	Q9FXB5 arabidopsis
26	43	53.1	643	10	Q39846	Q39846 glycine max
27	43	53.1	1011	11	Q9NMK1	Q9NMK1 mus musc
28	43	53.1	1131	5	Q9VGE4	Q9VGE4 dirosophila
29	43	53.1	1219	5	Q9VYJ9	Q9VYJ9 caenorhabdi
30	43	53.1	1645	5	Q9U263	Q9U263 caenorhabdi
31	43	53.1	1655	5	Q44498	Q44498 caenorhabdi
32	43	53.1	2785	11	Q99MW3	Q99MW3 mus musc
33	43	53.1	3259	4	Q14789	Q14789 homo sapien
34	42.5	52.5	670	4	Q9NMW3	Q9NMW3 homo sapien
35	42.5	52.5	682	4	Q9NX83	Q9NX83 homo sapien
36	42.5	52.5	746	4	Q9HA51	Q9HA51 homo sapien
37	42.5	52.5	975	4	Q9H9Y7	Q9H9Y7 homo sapien
38	42	51.9	251	11	Q9C8R1	Q9C8R1 mus musc
39	42	51.9	254	6	Q9GM10	Q9GM10 macaca fasc
40	42	51.9	419	3	Q74786	Q74786 schizosacch
41	42	51.9	438	6	Q9BH01	Q9BH01 macaca fasc
42	42	51.9	488	5	Q9GPA2	Q9GPA2 caenorhabdi
43	42	51.9	593	10	Q9MAB4	Q9MAB4 arabidopsis
44	42	51.9	726	2	Q9E0Z8	Q9E0Z8 escherichia
45	42	51.9	727	2	Q9E0Z4	Q9E0Z4 escherichia

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	321-AA
1	Q9CWB82	Q9CWB82		
AC	Q9CWB82	PRELIMINARY	PRT	321-AA
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DE	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	0610042106R1K PROTEIN (FRAGMENT).			
GN	0610042106R1K.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=KIDNEY;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Buft C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,			
RA	Hayashizaki Y.,			
RT	"Functional annotation of a full-length mouse cDNA collection."			
RL	Nature 409:685-690(2001).			
DR	EMBL: AK002914; BAB22453.1;			
DR	MGJ: 1514102; 0610042106R1K.			
DR	InterPro: IPR000719; Euk_kinase.			
DR	InterPro: IPR002290; Ser_thr_kin_actsite.			
DR	Pfam: PF00069; pkinase; I.			

DR SMART; SM00220; S\_TKC; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW ATP-binding; Transferase.  
 FT NON\_TER  
 SQ SEQUENCE 321 AA; 37201 MW; 2E96FDAE919F53A4 CRC64;

Query Match 95.1%; Score 77; DB 11; Length 321;  
 Best Local Similarity 93.8%; Pred. No. 0.0018;  
 Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIKAKRHDEOQRELE 16  
 DB 120 EIKAKRHDEOQRELE 135

RESULT 2  
 060877 PRELIMINARY; PRT; 445 AA.

DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DE PROTEIN\_KINASE\_MESS1.  
 GN STK3 OR MESS1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LIVER;  
 RA Han J.  
 RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U28726; AAA75300.1; -.  
 DR HSP; P00518; IPHK.  
 DR MGD; MGI:1928487; STK3.  
 DR InterPro: IPR000719; Euk. kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
 DR Pfam: PF00069; kinase; 1.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW ATP-binding; Transferase.  
 SQ SEQUENCE 445 AA; 50407 MW; 7B936B28DA616F05 CRC64;

Query Match 95.1%; Score 77; DB 11; Length 445;  
 Best Local Similarity 93.8%; Pred. No. 0.0024;  
 Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIKAKRHDEOQRELE 16  
 DB 296 EIKAKRHDEOQRELE 311

RESULT 3  
 09J110 PRELIMINARY; PRT; 497 AA.

AC 09J110;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE STE20-LIKE KINASE MST2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C;  
 RA Lee K.-K., Ohyama T., Yajima N., Yonehara S.;  
 FT "MST, a physiological caspase substrate, highly sensitizes apoptosis

RT both upstream and downstream of caspase activation";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF271361; AAF75790.1; -.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
 DR InterPro: IPR001245; Tyr\_kin.  
 DR Pfam; PF00069; kinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW ATP-binding; Kinase; Transferase.  
 SQ SEQUENCE 497 AA; 56855 MW; 9CDD365437581665 CRC64;

Query Match 95.1%; Score 77; DB 11; Length 497;  
 Best Local Similarity 93.8%; Pred. No. 0.0026;  
 Matches 15; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIKAKRHDEOQRELE 16  
 DB 296 EIKAKRHDEOQRELE 311

RESULT 4  
 054748 PRELIMINARY; PRT; 491 AA.

AC 054748;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DE MST2 KINASE.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Aurisicchio L., Dilauro R., Zannini M.;  
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ001529; CAA04814.1; -.  
 DR HSP; P00518; IPHK.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
 DR Pfam; PF00069; kinase; 1.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW ATP-binding; Transferase.  
 SQ SEQUENCE 491 AA; 56121 MW; 90FA6B020E7FFD62 CRC64;

Query Match 91.4%; Score 74; DB 11; Length 491;  
 Best Local Similarity 87.5%; Pred. No. 0.0069;  
 Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 EIKAKRHDEOQRELE 16  
 DB 296 EIKAKRHDEOQRELE 311

RESULT 5  
 060302 PRELIMINARY; PRT; 799 AA.

AC 060302;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE KIA0555 PROTEIN.  
 GN KIA0555.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:04:58 ; Search time 26.93 Seconds  
(without alignments)  
13.370 Million cell updates/sec

Title: US-09-822-110-3  
Perfect score: 81  
Sequence: 1 EIKAKRHDEQRELE 16

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued\_Patents\_AA:\*

1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfills1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	58.0	487	2	US-08-712-709-8
2	47	58.0	487	3	US-09-111-444-8
3	47	58.0	487	4	US-09-541-228-8
4	41	50.6	180	3	US-08-784-582-56
5	41	50.6	180	3	US-08-784-582-58
6	41	50.6	180	3	US-08-784-582-61
7	41	50.6	360	3	US-08-784-582-73
8	41	50.6	737	1	US-08-188-582-16
9	41	50.6	737	1	US-08-646-715-16
10	40	49.4	643	2	US-08-551-356-4
11	40	49.4	643	5	PCT-US93-12687-4
12	40	49.4	644	1	US-08-206-176-2
13	38.5	47.5	539	2	US-08-978-182-3
14	38.5	47.5	539	2	US-09-205-681-3
15	38	46.9	186	4	US-09-446-504-70
16	38	46.9	450	2	US-08-665-037-2
17	38	46.9	450	2	US-08-666-067-2
18	38	46.9	450	2	US-08-732-870-2
19	38	46.9	926	3	US-08-755-587-187
20	38	46.9	1964	3	US-08-836-325-10
21	38	46.9	1989	3	US-08-836-325-12
22	38	46.9	2548	4	US-09-172-422-1
23	37	45.7	35	4	US-09-248-588-26
24	37	45.7	203	1	US-08-216-593-5
25	37	45.7	203	5	PCT-US93-12380-5
26	37	45.7	205	1	US-07-992-827D-5
27	37	45.7	209	1	US-07-992-827D-2

28	37	45.7	209	1	US-08-216-593-2	Sequence 2, Appl1
29	37	45.7	209	5	PCT-US93-12380-2	Sequence 2, Appl1
30	37	45.7	288	3	US-08-312-949-4	Sequence 4, Appl1
31	37	45.7	288	3	US-08-446-201-4	Sequence 4, Appl1
32	37	45.7	289	1	US-08-072-070-4	Sequence 4, Appl1
33	37	45.7	289	1	US-08-469-434-4	Sequence 4, Appl1
34	37	45.7	289	1	US-08-214-222-5	Sequence 4, Appl1
35	37	45.7	289	2	US-08-467-852A-5	Sequence 5, Appl1
36	37	45.7	289	2	US-08-247-491A-5	Sequence 5, Appl1
37	37	45.7	289	2	US-07-992-827D-1	Sequence 1, Appl1
38	37	45.7	323	1	US-08-216-593-1	Sequence 1, Appl1
39	37	45.7	323	5	PCT-US93-12380-1	Sequence 1, Appl1
40	37	45.7	329	4	US-08-884-681-3	Sequence 3, Appl1
41	37	45.7	329	4	US-09-258-643-3	Sequence 3, Appl1
42	37	45.7	372	4	US-08-884-681-1	Sequence 1, Appl1
43	37	45.7	372	4	US-09-258-643-1	Sequence 1, Appl1
44	37	45.7	619	1	US-08-465-746-2	Sequence 2, Appl1
45	37	45.7				

## ALIGNMENTS

RESULT 1  
US-08-712-709-8  
; Sequence 8, Application US/08712709  
; Patent No. 5863780  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Hawkins, Phillip R.  
; TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/712,709  
; FILING DATE: Filed Herewith  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0118 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 487 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1117791  
; US-08-712-709-8

Query Match 58.0%; Score 47; DB 2; Length 487;  
Best Local Similarity 43.8%; Pred. No. 7.3;  
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 1 EIKAKRHDEQRELE 16

Db 299 DVKLKROESQOREMDQ 314

## RESULT 2

US-09-111-444-8

Sequence 8, Application US/09111444

Patent No. 6045792

GENERAL INFORMATION:

APPLICANT: Au-Young, Janice

APPLICANT: Guegler, Karl J.

APPLICANT: Hawkins, Phillip R.

TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: U.S.

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/111,444

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/712,709

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0118 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 487 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 1117791

US-09-111-444-8

Query Match 58.0%; Score 47; DB 3; Length 487;

Best Local Similarity 43.8%; Pred. No. 7.3;

Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 EIKAKRHDEQORELEE 16

Db 299 DVKLKROESQOREMDQ 314

## RESULT 3

US-09-541-228-8

Sequence 8, Application US/09541228

Patent No. 6232077

GENERAL INFORMATION:

APPLICANT: Au-Young, Janice

APPLICANT: Guegler, Karl J.

APPLICANT: Hawkins, Phillip R.

TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: U.S.

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/541,228

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/712,709

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0118 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 487 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 1117791

US-09-541-228-8

Query Match 58.0%; Score 47; DB 4; Length 487;

Best Local Similarity 43.8%; Pred. No. 7.3;

Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 EIKAKRHDEQORELEE 16

Db 299 DVKLKROESQOREMDQ 314

## RESULT 4

US-08-784-582-56

Sequence 56, Application US/08784582

Patent No. 6110707

GENERAL INFORMATION:

APPLICANT: Newgard, Christopher B.

APPLICANT: Halban, Philippe A.

APPLICANT: No. 6110707Mington, Karl D.

APPLICANT: Clark, Samuel A.

APPLICANT: Thijsen, Anice E.

APPLICANT: Quade, Christian

APPLICANT: Kruse, Fred

APPLICANT: McGarry, Dennis

TITLE OF INVENTION: RECOMBINANT EXPRESSION OF PROTEINS FROM

SECRETORY CELL LINES

NUMBER OF SEQUENCES: 79

CORRESPONDENCE ADDRESS:

ADDRESS: Arnold, White &amp; Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:05:59 ; Search time 54.15 seconds  
(without alignments)  
21.887 Million cell updates/sec

Title: US-09-822-110-3

Perfect score: 81  
Sequence: 1 EIKARKHDEQORELEE 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A\_Geneseq\_1101:\*

1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT:\*

2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT:\*

3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT:\*

4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT:\*

5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT:\*

6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT:\*

7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT:\*

8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT:\*

9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT:\*

10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT:\*

11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT:\*

12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT:\*

13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT:\*

14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT:\*

15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT:\*

16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT:\*

17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT:\*

18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT:\*

19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT:\*

20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT:\*

21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT:\*

22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted, by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	59.3	62	22	AA18700
2	45	55.6	487	20	AA121674
3	45	55.6	625	19	AAW62830
4	45	55.6	666	19	AAW62828
5	45	55.6	666	19	AAW62829
6	45	55.6	1392	20	AAW66999
7	45	55.6	1427	12	AA10534
8	43	53.1	88	21	AA622744
9	43	53.1	100	21	AA622743
10	43	53.1	363	22	AA893034
11	42.5	52.5	243	21	AA840619

12	42.5	52.5	746	22	AA893938	Human protein sequ
13	42.5	52.5	975	22	AA894042	Human protein sequ
14	42	51.9	164	22	AA863548	Human gastric cance
15	42	51.9	283	22	AA673799	Human colon cancer
16	42	51.9	482	20	AA07067	Renal cancer assoc
17	42	51.9	752	18	AAW34178	Human transcriptio
18	42	51.9	898	20	AAV55936	Human SUI1 protei
19	42	51.9	898	20	AAV97676	Human KDS1 serine/
20	42	51.9	898	21	AA444244	Human cell signall
21	42	51.9	899	21	AA843191	Human ORFX ORF2955
22	42	51.9	1299	22	AA893286	Human protein sequ
23	42	51.9	1299	21	AAV58653	Protein regulating
24	41	50.6	121	22	AA838088	S. epidermidis ope
25	41	50.6	156	22	AAU02907	Angiotensin conver
26	41	50.6	180	18	AAW22079	Rat preproglucagon
27	41	50.6	180	18	AAW22080	Human preproglucag
28	41	50.6	180	18	AAW22081	Human preproglucag
29	41	50.6	180	18	AAW16384	Rat prepro-glucago
30	41	50.6	180	20	AAV39812	Preproglucagon pro
31	41	50.6	180	21	AA826773	Human preproglucag
32	41	50.6	180	21	AA826774	Human preproglucag
33	41	50.6	180	21	AA826775	Mutant human prepr
34	41	50.6	360	21	AA826777	Human growth hormo
35	41	50.6	411	22	AAW39744	Human polypeptide
36	41	50.6	430	22	AAW41530	Human polypeptide
37	41	50.6	478	22	AA670772	S cerevisiae apopt
38	41	50.6	737	15	AA856494	TARA-binding prote
39	41	50.6	737	17	AAW06084	Human TARA-binding
40	41	50.6	737	18	AAW25019	TARA-binding prote
41	40.5	50.0	332	16	AAW04952	Leucocytozoan prot
42	40.5	50.0	1562	16	AAW04951	Leucocytozoan prot
43	40	49.4	271	22	AA880637	Environmental stre
44	40	49.4	360	21	AA854135	Human pancreatic c
45	40	49.4	643	15	AA860020	Fibronectin. Homo

# ALIGNMENTS

RESULT 1	AA18700	standard; Protein: 62 AA.
ID	AA18700	
AC	AA18700	
DT	12-OCT-2001	(first entry)
DE	Peptide #5134 encoded by probe for measuring cervical gene expression.	
XX	Probe: human; microarray; gene expression; cervical epithelial cell;	
KW	cervical cancer.	
XX		
OS	Homo sapiens.	
XX		
PN	W0200157278-A2.	
PD	09-AUG-2001.	
PF	30-JAN-2001; 2001WO-US00670.	
XX		
PR	04-FEB-2000; 2000US-0180312.	
PR	26-MAY-2000; 2000US-0207456.	
PR	30-JUN-2000; 2000US-0608408.	
PR	03-AUG-2000; 2000US-0632366.	
PR	21-SEP-2000; 2000US-0234687.	
PR	27-SEP-2000; 2000US-0236359.	
PR	04-OCT-2000; 2000GB-0024263.	
XX		
PA	(MOLE-) MOLECULAR DYNAMICS INC.	
XX		
PI	Penn SG, Hanzel DK, Chen W, Rank DR;	
XX		
DR	WPI; 2001-488901/53.	

XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells -  
 PS  
 XX Claim 27; SEQ ID No 23526; 487bp; English.  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SEMP: see AAI10068-AI128459). The present sequence is a peptide encoded  
 CC by one such probe. The SEMP is derived from human Hela cells. The SEMP  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 62 AA;  
 OY 1 EIKAKRHDEQORELE 15  
 DB 29 elkakheekkelq 43  
 RESULT 2  
 AAY21674 standard; Protein; 487 AA.  
 XX ID AAY21674;  
 AC AAY21674;  
 XX 18-AUG-1999 (first entry)  
 DT  
 XX Human Ste20 homologue polypeptide S201.  
 DE  
 XX Sterile 20; serine/threonine kinase; C12.2bs; 5e.new; murine;  
 KW erythroleukemia; mitogenic signaling; cell differentiation; metastasis;  
 KW signaling pathway; cancer; ischemic stroke; heart disease; inflammation.  
 XX  
 XX Homo sapiens.  
 OS  
 XX MO9929857-A1.  
 XX 17-JUN-1999.  
 XX 09-DEC-1998; 98WO-US26116.  
 PF  
 XX 09-DEC-1997; 97US-0069078.  
 PR  
 XX (CHIL-) CHILDRENS MEDICAL CENT.  
 XX  
 XX Agarwal S, Best J, Vail B, Zon LI;  
 PI WPI; 1999-385605/32.  
 DR  
 XX Murine homologues of the sterile 20 family of serine/threonine  
 PT kinases, useful in elucidation of mitogenic signalling pathways  
 PT  
 XX Disclosure; Fig 4; 37pp; English.  
 PS  
 XX The invention provides two members of the sterile 20 (Ste20) family of  
 CC serine/threonine kinases (C12.2bs and 5e.new). Host cells transfected  
 CC with vectors comprising the nucleic acids encoding the polypeptides are  
 CC used for their recombinant expression. The Ste20 kinases, isolated from a  
 CC murine erythroleukemia cDNA library, are regulatory molecules involved in  
 CC mitogenic signaling as well as other cellular phenomena such as  
 CC morphology and motility. These activities are important factors in  
 CC development, cell differentiation, cancer and metastases. Manipulation of

CC the C12.2bs and 5e.new proteins or polynucleotides are useful for  
 CC manipulation of the signalling pathways involved and will allow  
 CC development of reagents to modulate these signalling pathways. The kinases  
 CC are also important for ischemic stroke, heart disease and inflammation.  
 CC The antibodies can be used to inhibit the activity of the kinase,  
 CC especially in vitro and in cell extracts, as well as identify the kinases  
 CC in immunoassays. The present sequence represents a S201 polypeptide, a  
 CC human homologue of Ste20.  
 CC  
 XX Sequence 487 AA;  
 OY 1 EIKAKRHDEQORELE 16  
 DB 299 dvklkrqeqagrevdq 314  
 RESULT 3  
 AAW62830 standard; Protein; 625 AA.  
 XX ID AAW62830;  
 AC AAW62830;  
 XX 27-OCT-1998 (first entry)  
 DT  
 XX Macadamia integrifolia antimicrobial protein.  
 DE  
 XX antimicrobial protein; infestation; control.  
 KW  
 XX Macadamia integrifolia.  
 OS  
 XX Key Location/Qualifiers  
 FH Peptide 1..28  
 FT /note="signal peptide"  
 FT 29..666  
 FT Protein /note="mature protein"  
 FT  
 XX WO9827805-A1.  
 PN  
 XX 02-JUL-1998.  
 PD  
 XX 22-DEC-1997; 97WO-AU00874.  
 PF  
 XX 20-DEC-1996; 96AU-0004275.  
 PR  
 XX (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY;  
 PA Bower NI, Goulier KC, Green JL, Manners JM, Marcus JP;  
 XX WPI; 1998-377279/32.  
 PI N-PSDB; AAV42316.  
 DR  
 XX Novel anti-microbial protein from e.g. Macadamia integrifolia -  
 XX useful for controlling microbial infestations of plants or mammals  
 PT  
 XX Claim 1; Page 43-45; 96pp; English.  
 PS  
 XX The sequence is that of an antimicrobial protein which can  
 CC be used to control microbial infestations in plants and mammalian  
 CC animals.  
 CC  
 XX Sequence 625 AA;  
 OY 5 KRHDEQORELE 16  
 Query Match 55.6%; Score 45; DB 19; Length 625;  
 Best Local Similarity 75.0%; Pred. No. 45;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OW protein - protein search, using sw model

Run on: February 15, 2002, 07:08:32 ; Search time 12.48 Seconds  
(without alignments)  
28.850 Million cell updates/sec

Title: US-09-822-110-3

Sequence: 1 EIKAKRHDEQRELE 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

1 number of hits satisfying chosen parameters: 96934

Minimum DB seq length: 0  
Maximum DB seq length: 16

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	34.6	12	2	US-08-973-563A-29
2	28	34.6	12	2	US-08-973-559-29
3	28	34.6	12	2	US-08-973-235-9
4	28	34.6	12	4	US-09-051-986-12
5	28	34.6	14	1	US-07-686-322A-6
6	28	34.6	14	1	US-08-002-999-6
7	28	34.6	14	4	US-09-051-986-13
8	27	33.3	12	4	US-08-199-776-24
9	27	33.3	12	2	US-08-973-563A-31
10	27	33.3	12	2	US-08-973-559-31
11	27	33.3	12	3	US-08-663-731-24
12	27	33.3	12	3	US-08-663-731-24
13	27	33.3	12	3	US-08-663-731-24
14	27	33.3	12	3	US-08-663-731-24
15	27	33.3	12	5	PCT-US95-02044-24
16	27	33.3	13	4	US-09-331-507A-10
17	27	33.3	13	4	US-09-331-507A-25
18	27	33.3	14	1	US-08-199-776-18
19	27	33.3	14	1	US-08-352-179-27
20	27	33.3	14	3	US-08-663-731-18
21	27	33.3	14	5	PCT-US95-02044-18
22	27	33.3	15	1	US-08-199-776-20
23	27	33.3	15	3	US-08-663-731-20
24	27	33.3	15	3	US-08-663-731-20
25	27	33.3	15	5	PCT-US95-02044-20
26	27	33.3	16	1	US-08-126-564A-36
27	27	33.3	16	5	PCT-US94-09143-36

28	26	32.1	11	6	5210075-10	Patent No. 5210075
29	26	32.1	12	4	US-09-518-046-34	Sequence 34, Appl
30	26	32.1	13	1	US-07-987-272A-6	Sequence 6, Appl
31	26	32.1	14	1	US-08-199-776-19	Sequence 19, Appl
32	26	32.1	14	3	US-08-663-731-19	Sequence 19, Appl
33	26	32.1	14	3	US-08-663-731-19	Sequence 19, Appl
34	26	32.1	14	5	PCT-US95-02044-19	Sequence 19, Appl
35	26	32.1	15	1	US-08-287-717-7	Sequence 7, Appl
36	26	32.1	15	1	US-08-481-888A-17	Sequence 17, Appl
37	26	32.1	15	1	US-08-485-273A-17	Sequence 17, Appl
38	26	32.1	15	1	US-08-441-914-7	Sequence 7, Appl
39	26	32.1	15	2	US-08-973-563A-17	Sequence 17, Appl
40	26	32.1	15	2	US-08-973-559-17	Sequence 17, Appl
41	25	30.9	9	5	PCT-US95-03236-56	Sequence 56, Appl
42	25	30.9	10	1	US-08-481-888A-21	Sequence 21, Appl
43	25	30.9	10	1	US-08-485-273A-21	Sequence 21, Appl
44	25	30.9	10	2	US-08-973-563A-21	Sequence 21, Appl
45	25	30.9	10	2	US-08-973-559-21	Sequence 21, Appl

## ALIGNMENTS

RESULT 1  
US-08-973-563A-29  
Sequence 29, Application US/08973563A  
Patent No. 5885965  
GENERAL INFORMATION:  
APPLICANT: Oppenheim, Frank G.  
APPLICANT: Xu, Tao  
APPLICANT: Spacciapoli, Peter  
APPLICANT: Roberts, F. D.  
APPLICANT: Friden, Philip M.  
TITLE OF INVENTION: Anti-Fungal D-Amino Acid Histatin-Based  
Peptides  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millita Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973,563A  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/485,273  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: PER95-02A2  
TELEPHONE: 781-861-6240  
TELEFAX: 781-861-9540  
INFORMATION FOR SEQ. ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..12  
OTHER INFORMATION: /note= "At least one amino acid

OTHER INFORMATION: must have a D configuration."  
US-08-973-563A-29

Query Match 34.6%; Score 28; DB 2; Length 12;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 AKRHDEQORE 13  
|||||:|:  
DB 1 AKRHHKRYRK 10

RESULT 2  
US-08-973-559-29  
Sequence 29, Application US/08973559  
Patent No. 5912230  
GENERAL INFORMATION:  
APPLICANT: OPPENHEIM, FRANK G.  
APPLICANT: XU, TAO  
APPLICANT: ROBERTS, F. D.  
APPLICANT: SPACCIAPOLI, PETER  
APPLICANT: FRIDEN, PHILIP M.  
TITLE OF INVENTION: Anti-Fungal and Anti-Bacterial  
TITLE OF INVENTION: Histatin-Based Peptides  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millita Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: US  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973,559  
FILING DATE: 07-JUN-1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/481,888  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: PER95-01A2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 781-861-6240  
TELEFAX: 781-861-9540  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-973-559-29

Query Match 34.6%; Score 28; DB 2; Length 12;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 AKRHDEQORE 13  
|||||:|:  
DB 1 AKRHHKRYRK 10

RESULT 3  
US-08-993-235-9

Sequence 9, Application US/08993235  
Patent No. 6084064  
GENERAL INFORMATION:  
APPLICANT: OPPENHEIM, FRANK G.  
APPLICANT: ROBERTS, F. DONALD  
APPLICANT: XU, TAO  
APPLICANT: SPACCIAPOLI, PETER  
TITLE OF INVENTION: ANTIFUNGAL AND ANTIBACTERIAL PEPTIDES  
FILE REFERENCE: 50032/002001  
CURRENT APPLICATION NUMBER: US/08/993,235  
CURRENT FILING DATE: 1997-12-18  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 9  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-08-993-235-9

Query Match 34.6%; Score 28; DB 3; Length 12;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 AKRHDEQORE 13  
|||||:|:  
DB 1 AKRHHKRYRK 10

RESULT 4  
US-09-051-986-12  
Sequence 12, Application US/09051986A  
Patent No. 6191113  
GENERAL INFORMATION:  
APPLICANT: NAKAHARA, YO  
APPLICANT: HARA, Saburo  
APPLICANT: KAMIKUBO, Yuichi  
APPLICANT: TAKEMOTO, Sumiyo  
APPLICANT: MIYAMOTO, Seiji  
TITLE OF INVENTION: NOVEL PEPTIDE  
FILE REFERENCE: NAKAHARA-1  
CURRENT APPLICATION NUMBER: US/09/051,986A  
CURRENT FILING DATE: 1998-04-24  
EARLIER APPLICATION NUMBER: JP 300792/1995  
EARLIER FILING DATE: 1995-10-24  
EARLIER APPLICATION NUMBER: PCT/JP96/03080  
EARLIER FILING DATE: 1996-10-23  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 12  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Human Tissue Factor Pathway Inhibitor  
US-09-051-986-12

Query Match 34.6%; Score 28; DB 4; Length 12;  
Best Local Similarity 54.5%; Pred. No. 1.3e+02;  
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 IKAKRHDEQOR 12  
|||:|:  
DB 1 IKTKRRRRKOR 11

RESULT 5  
US-07-686-322A-6  
Sequence 6, Application US/07686322A  
Patent No. 5312733  
GENERAL INFORMATION:  
APPLICANT: MacLeod Dr., Carol L.  
TITLE OF INVENTION: NO. 5312733el T-cell Lymphoma cDNA Clones  
NUMBER OF SEQUENCES: 6



PI Rath M;  
 XX  
 DR WPI; 1995-263953/34.  
 XX  
 PT Identifying signal oligopeptide(s) in protein sequence(s) - shown as  
 PT regions of max. hydrophilicity, used in modulating communication  
 PT between protein(s)  
 PS  
 XX  
 PS Claim 5; Page 44; 88pp; English.  
 CC The sequences given in AAM21201-560 represent hydrophilic signal oligo-  
 CC peptides. These signal oligopeptides are localised on the surface  
 CC of the protein and are represented by the hydrophilicity maxima of  
 CC the protein. These peptides are enriched in charged amino acids  
 CC arranged with neutral spacer amino acids. The specific signal  
 CC character of these oligopeptides is determined by a characteristic  
 CC combination of conformation and charge within the signal sequence.  
 CC These oligopeptides may be used as vaccines in the treatment of  
 CC human disease, as competitive inhibitors to prevent or reduce the  
 CC metabolic action or interaction of a selected protein by blocking  
 CC its specific signal sequences, or as therapeutic agents to function  
 CC as feedback regulators to reduce synthesis rate of a selected protein.  
 CC These peptides may be modified by omitting one or more amino acids at  
 CC the N- and/or C-terminal, by substituting one or more amino acids  
 CC without consideration of charge and polarity, by substituting one or  
 CC more amino acids with amino acid residues with similar charge and/or  
 CC polarity, by omitting one or more amino acids or a combination of these.  
 CC  
 XX  
 SQ Sequence 9 AA:  
 Query Match 40.7%; Score 33; DB 16; Length 9;  
 Best Local Similarity 75.0%; Pred. No. 4.3e+05;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 5 KRHDEQOR 12  
 Db 1 krhdefer 8  
 RESULT 2  
 AAG98093  
 ID AAG98093 standard; Peptide; 14 AA.  
 AC AAG98093;  
 DT 19-SEP-2001 (first entry)  
 PT Human SNP associated peptide SEQ ID NO. 735.  
 XX  
 KW Human; single nucleotide polymorphism; SNP; angiotensin;  
 KW 4-hydroxybutyrate; dehydrogenase; protein therapy;  
 KW adenosine triphosphate-dependent RNA helicase;  
 KW major histocompatibility complex Class I histocompatibility antigen; MHC;  
 KW phosphoglycerate kinase; immunosuppressive; immunostimulatory;  
 KW antineumatic; antisclerotic; antidiabetic; antiinflammatory; cytostatic;  
 KW antileukemic; neuroprotective; antimicrobial; gene therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200148245-A2.  
 PD 05-JUL-2001.  
 XX  
 PF 27-DEC-2000; 2000WO-US53346.  
 XX  
 PR 27-DEC-1999; 99US-0472688.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shinkets RA, Leach M;  
 XX  
 DR WPI; 200A-418297/44.

XX  
 PT Polymorphic nucleic acids encoding e.g. angiotensin, dehydrogenase,  
 PT adenosine triphosphate-dependent RNA helicase and/or phosphoglycerate  
 PT kinase, useful for diagnosing and treating, e.g. cancer, autoimmune  
 PT diseases and infections -  
 PT  
 PS Disclosure; Page 452; 484pp; English.  
 XX  
 CC The invention relates to nucleic acids (AAH79386-AAH80036) encoding  
 CC polymorphic variants of proteins (AAG8010-AAG8238) related to  
 CC angiotensin, 4-hydroxybutyrate, dehydrogenase, adenosine triphosphate  
 CC (ATP)-dependent RNA helicase, major histocompatibility complex (MHC)  
 CC Class I histocompatibility antigen and/or phosphoglycerate kinase. These  
 CC nucleic acid single nucleotide polymorphisms (SNPs) and the encoded  
 CC proteins have potential immunosuppressive, immunostimulatory,  
 CC antineumatic, antisclerotic, antidiabetic, antiinflammatory, cytostatic,  
 CC antileukemic, neuroprotective and antimicrobial activity and may be  
 CC useful in gene/protein therapy, vaccines, modulation of the expression  
 CC and activity of proteins related to angiotensin, 4-hydroxybutyrate,  
 CC dehydrogenase, adenosine triphosphate (ATP)-dependent RNA helicase,  
 CC major histocompatibility complex (MHC) Class I histocompatibility antigen  
 CC and/or phosphoglycerate kinase. Disorders that may be prevented,  
 CC diagnosed and/or treated by the above methods include multifactorial  
 CC diseases with a genetic component, such as autoimmune diseases (e.g.  
 CC rheumatoid arthritis, multiple sclerosis, diabetes, systemic lupus  
 CC erythematosus and Grave's disease), inflammation, cancer (e.g. cancers  
 CC of the bladder, brain, breast, colon and kidney, leukemia), diseases of  
 CC the nervous system, an infection of pathogenic organisms. They may also  
 CC be used to alter phenotypic traits such as longevity, appearance,  
 CC strength, speed and endurance.  
 CC  
 XX  
 SQ Sequence 14 AA:  
 Query Match 40.7%; Score 33; DB 22; Length 14;  
 Best Local Similarity 46.2%; Pred. No. 52;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 EIRAKRHDEQORE 13  
 Db 1 eqqeqnqqeqq 13  
 RESULT 3  
 AAM21333  
 ID AAM21333 standard; peptide; 11 AA.  
 AC AAM21333;  
 DT 29-JUL-1997 (first entry)  
 PT Glucagon precursor derived signal oligopeptide #38.  
 XX  
 DE  
 KW Hydrophilic; signal oligopeptide; hydrophilicity maxima; vaccine; STV;  
 KW competitive inhibitor; feedback regulator; synthesis; gastrin precursor;  
 KW charge; polarity; farnesyl synthetase; plasminogen activator inhibitor 1;  
 KW hydroxymethylglutaryl coenzyme A reductase; glucagon precursor; rheus;  
 KW gonadoliberin precursor; plasminogen activator inhibitor 2; protein;  
 KW Alzheimer amyloid A4; corticotropin releasing factor binding protein;  
 KW apolipoprotein E; herpes virus 1 glycoprotein B; HSV1; human; OMVVS;  
 KW Treponema pallidum membrane protein; TMAP; islet amyloid polypeptide;  
 KW fibroblast MMP1; schistosoma elastase precursor; schistosomins;  
 KW hepatitis delta antigen; rev protein; HIV; VILV; angiotensinogen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9519568-A1.  
 XX  
 PD 20-JUL-1995.  
 XX  
 PF 12-JAN-1995; 95WO-US00575.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:15:07 ; Search time 22.13 seconds  
(without alignments)  
105.755 Million cell updates/sec

Title: US-09-822-110-3

Perfect score: 81  
Sequence: 1 EIKAKRHDEQRELEE 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

al number of hits satisfying chosen parameters: 3030

Minimum DB seq length: 0

Maximum DB seq length: 16

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhcc:\*  
8: sp.organella:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.podent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	33.3	16	4	Q9UC12
2	25	30.9	7	6	Q28742
3	25	30.9	14	10	P82333
4	24	29.6	13	12	O9OCL8
5	24	29.6	16	11	O9OUW8
6	23	28.4	15	5	Q9TW74
7	23	28.4	16	6	P79137
8	22	27.2	15	4	O9UCN2
9	22	27.2	16	4	O9UD21
10	22	27.2	16	6	O9NDF6
11	22	27.2	16	6	O9NDF6
12	21	25.9	13	8	O9T569
13	21	25.9	16	4	O16413
14	20	24.7	13	4	O14804
15	20	24.7	15	9	O9UCO0
16	20	24.7	15	9	O38574
17	20	24.7	16	2	O9RP22
18	19	23.5	9	6	O9TW77
19	19	23.5	11	10	O82070

20	19	23.5	11	10	P82336	P82336 pisum sativ
21	19	23.5	13	3	O9URU0	O9URU0 schizosacch
22	19	23.5	13	12	O9OCM4	O9OCM4 borra disea
23	19	23.5	13	12	O9OCM1	O9OCM1 borra disea
24	19	23.5	13	12	O9OCL5	O9OCL5 borra disea
25	19	23.5	13	12	O9OCL2	O9OCL2 borra disea
26	19	23.5	13	12	O9OCK9	O9OCK9 borra disea
27	19	23.5	13	12	O9OCK6	O9OCK6 borra disea
28	19	23.5	13	12	O9OCK3	O9OCK3 borra disea
29	19	23.5	14	2	O47599	O47599 escherichia
30	19	23.5	14	2	O9KCS9	O9KCS9 bacillus ha
31	19	23.5	14	12	O56127	O56127 porcine cit
32	19	23.5	15	3	O9URE0	O9URE0 saccharomyc
33	19	23.5	15	3	O9UR93	O9UR93 candida alb
34	19	23.5	15	6	O9T009	O9T009 bos taurus
35	19	23.5	15	6	O9TR40	O9TR40 bos taurus
36	19	23.5	15	10	P82936	P82936 hordium vul
37	19	23.5	16	4	O9NY32	O9NY32 homo sapien
38	19	23.5	16	8	O9T206	O9T206 solanum tub
39	18	22.2	8	12	O84273	O84273 human papill
40	18	22.2	10	11	O9OVE8	O9OVE8 mus sp. pro
41	18	22.2	12	11	O9OVE2	O9OVE2 rattus sp.
42	18	22.2	13	10	O9S8N1	O9S8N1 hordium vul
43	18	22.2	14	5	O26075	O26075 psamechinu
44	18	22.2	14	11	O9UJ05	O9UJ05 mus musculu
45	18	22.2	14	11	O9CS58	O9CS58 mus musculu

#### ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	16 AA.
ID: O9UC12				
AC: O9UC12				
DT: 01-MAY-2000 (TREMBlrel. 13, Created)				
DT: 01-MAY-2000 (TREMBlrel. 13, Last sequence update)				
DT: 01-JUN-2001 (TREMBlrel. 17, Last annotation update)				
DE: TROPOMYOSIN (FRAGMENT)				
OS: Homo sapiens (Human)				
OC: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC: Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo;				
OX: NCBI_TaxId=9606;				
RN: [1]				
RP: SEQUENCE.				
RF: MEDLINE=93193532; Pubmed=8450225;				
RA: Des K.M., Dasgupta A., Mandal A., Geng X.;				
RT: "Autoimmunity to cytoskeletal protein tropomyosin. A clue to the				
RT: pathogenetic mechanism for ulcerative colitis."				
RL: J. Immunol. 150:2487-2493(1993).				
DR: InterPro: IPR000533; Tropomyosin.				
DR: Pfam: PF00261; Tropomyosin; 1.				
SO: SEQUENCE	16 AA:	1960 MW;	7A26C4941E0A335E	CRC64;
Query Match	33.3%	Score 27;	DB 4;	Length 16;
Best Local Similarity	44.4%	Pred. No. 1.4e+03;		
Matches	4;	Conservative	4;	Mismatches 1;
				Indels 0;
				Gaps 0;
QY	8 DEQRELEE 16			
Db	11::1::1			
	4 DEKMEIQE 12			
RESULT	2			
ID: O28742				
AC: O28742		PRELIMINARY:	PRT:	7 AA.
DT: 01-NOV-1996 (TREMBlrel. 01, Created)				
DT: 01-NOV-1996 (TREMBlrel. 01, Last sequence update)				
DT: 01-NOV-1998 (TREMBlrel. 08, Last annotation update)				
DE: ALPHA-MYOSIN HEAVY CHAIN (FRAGMENT)				
OS: Oryctolagus cuniculus (Rabbit).				

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OC NCBI\_TaxID=9986;  
 RN [1]  
 RP MEDLINE=84221901; PubMed=6328491;  
 RA Friedman D.J., Umeda P.K., Sinha A.M., Hsu H.J., Jokovic S.,  
 RA Rabinowitz M.;  
 RT "Characterization of genomic clones specifying rabbit alpha- and beta-  
 RT ventricular myosin heavy chains."  
 RL Proc. Natl. Acad. Sci. U.S.A. 81:3044-3048(1984).  
 DR EMBL; K01698; AAA31415.1;  
 KW Myosin.  
 FT NON\_TER  
 SO SEQUENCE 7 AA; 916 MW; 6B1B1AA1E69326B0 CRC64;

QY 5 RHDEQ 10  
 Db 2 KMHDE 7

Query Match 30.9%; Score 25; DB 6; Length 7;  
 Best Local Similarity 66.7%; Pred. No. 4.7e+05;  
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 3  
 ID P82333 PRELIMINARY; PRT; 14 AA.  
 AC P82333;  
 DT 01-JUN-2000 (TREMBlrel. 14, Created)  
 DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
 DE UNKNOWN PROTEIN FROM 2D-PAGE OF THYLAKOID (SPOT118) (FRAGMENT).  
 OS Pisum sativum (Garden pea).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids I; Fabales; Fabaceae; Papilionoideae; Pisum.  
 OC NCBI\_TaxID=3888;  
 RN [1]  
 RP SEQUENCE, SUBCELLULAR LOCATION, AND DEVELOPMENTAL STAGE.  
 RC STRAIN=CV; DE GRACE; TISSUE=LEAF;  
 RX MEDLINE=20181728; PubMed=10715320;  
 RA Peltier J.-B., Friso G., Kalume D.E., Roepstorff P., Nilsson F.,  
 RA Adamska I., van Wijk K.J.;  
 RT "Proteomics of the chloroplast: systematic identification and  
 RT targeting analysis of luminal and peripheral thylakoid proteins."  
 RT Plant Cell 12:319-341(2000).  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE LUMEN OR  
 CC PERIPHERY.  
 CC -1- DEVELOPMENTAL STAGE: UNFOLDED AND FULLY DEVELOPED LEAVES.  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 5.7, ITS MW IS: 28 KDA.  
 KW Chloroplast; Thylakoid membrane.  
 FT NON\_TER  
 SO SEQUENCE 14 AA; 1723 MW; CD839374AAF2CA06 CRC64;

OY 8 DEORELE 16  
 Db 1 EEOREDO 9

Query Match 30.9%; Score 25; DB 10; Length 14;  
 Best Local Similarity 44.4%; Pred. No. 2.5e+03;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

RESULT 4  
 ID O9QC18 PRELIMINARY; PRT; 13 AA.  
 AC O9QC18;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE P40 ORF (FRAGMENT).  
 OC Borna disease virus (BDV).  
 OS Viruses; ssRNA negative-strand viruses; Mononegavirales.  
 OC NCBI\_TaxID=12455;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99445641; PubMed=10515835;  
 RA Czayan M., Hallensleben W., Hofer M., Pollak S., Sauder C., Bilzer T.,  
 RA Blumcke I., Riederer P., Bogerts B., Falkai P., Schwarz M.J.,  
 RA Masliah E., Staeheli P., Hufert F.T., Lieb K.;  
 RT "Borna disease virus in human brains with a rare form of hippocampal  
 RT degeneration but not in brains of patients with common  
 RT neuropsychiatric disorders."  
 RL J. Infect. Dis. 180:1695-1699(1999).  
 DR EMBL; AJ246860; CAB52093.1;  
 FT NON\_TER  
 FT NON\_TER  
 SO SEQUENCE 13 AA; 1673 MW; 8F80680E534AA0A CRC64;

QY 6 RHDEQ 12  
 Db 1 RHENDR 7

Query Match 29.6%; Score 24; DB 12; Length 13;  
 Best Local Similarity 57.1%; Pred. No. 3.2e+03;  
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RESULT 5  
 ID O9QW8 PRELIMINARY; PRT; 16 AA.  
 AC O9QW8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DE 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE PROTEIN ISOMERASE-RELATED PROTEIN PRECURSOR 71.5 KDA ISOFORM  
 DE (FRAGMENT).  
 OS Cavia (guinea pigs).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Cavidae.  
 OC NCBI\_TaxID=10140;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=96070119; PubMed=8535285;  
 RA Bonifacio M.D., Steeves T., Saunders D.M., Sinosich M.J.;  
 RT "Isolation of Erp72 from guinea pig term placenta using heparin  
 RT Sepharose affinity chromatography."  
 RL Biochem. Mol. Biol. Int. 36:1143-1152(1995).  
 SO SEQUENCE 16 AA; 1880 MW; 21B44A5F5767CB17 CRC64;

OY 8 DEORELE 15  
 Db 3 DEEDDLE 10

Query Match 29.6%; Score 24; DB 11; Length 16;  
 Best Local Similarity 50.0%; Pred. No. 3.8e+03;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

RESULT 6  
 ID O9TWT4 PRELIMINARY; PRT; 15 AA.  
 AC O9TWT4;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)  
 DE HEMOGLOBIN (FRAGMENT).  
 OS Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;  
 OC Lumbricina; Lumbricidae; Lumbricus.



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:15:22 ; Search time 10.02 Seconds

(without alignments)  
58.547 Million cell updates/sec

Title: US-09-822-110-3

Perfect score: 81

Sequence: 1 EIKAKRHDQORELEE 16

Scoring table: BLOSUM62

Searched: 100059 segs, 36664827 residues 780

Minimum DB seq length: 0  
Maximum DB seq length: 16

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	29.6	14	1	RS19_PRUAP
2	24	29.6	16	1	LE06_BIOGL
3	21	25.9	14	1	YMTF_XANCP
4	20	24.7	15	1	UN01_PINPS
5	19	23.5	9	1	NEUX_HUMAN
6	19	23.5	9	1	NEUX_RAT
7	19	23.5	12	1	RS19_TOBBP
8	19	23.5	13	1	PRC_SALTY
9	18	22.2	10	1	COX4_THOGB
10	17	21.0	10	1	FIBB_CERSI
11	17	21.0	14	1	SCK3_LEIOU
12	16	19.8	5	1	TRM3_ECOLI
13	16	19.8	7	1	UF03_MOUSE
14	16	19.8	12	1	RS19_CLEYE
15	16	19.8	12	1	XYLA_STEVN
16	16	19.8	13	1	NEUT_BUFMA
17	16	19.8	13	1	NEUT_CHICK
18	16	19.8	13	1	NEUT_RANTE
19	16	19.8	13	1	NEUT_TRIVU
20	16	19.8	14	1	DCM_PSECF
21	16	19.8	15	1	ACEM_ACICA
22	16	19.8	16	1	BALI_EUBSP
23	16	19.8	16	1	BL6_VIBRP
24	16	19.8	16	1	YMOR_PSEPU
25	15	18.5	10	1	MALE_KLEPN
26	15	18.5	10	1	SYK_CAMOP
27	15	18.5	13	1	NEUT_CAVPO
28	15	18.5	15	1	GR78_HORSE
29	15	18.5	15	1	UC27_MAIZE
30	15	18.5	15	1	UN04_PINPS
31	15	18.5	16	1	ALLI_CAIYO
32	15	18.5	16	1	LPHI_ECOLI
33	14	17.3	6	1	TRPI_PSEPU

34	14	17.3	8	1	GLUR_HUMAN
35	14	17.3	8	1	UH09_RAT
36	14	17.3	10	1	GLEM_HUMAN
37	14	17.3	10	1	GON3_PETMA
38	14	17.3	10	1	TRNK_PIG
39	14	17.3	10	1	UH05_RAT
40	14	17.3	11	1	CS15_BACSU
41	14	17.3	12	1	LICH_BACLI
42	14	17.3	12	1	PA2B_VIBPO
43	14	17.3	12	1	PPK4_PERFU
44	14	17.3	12	1	RR16_GIMBI
45	14	17.3	12	1	TA10_TREME

## ALIGNMENTS

RESULT 1	RS19_PRUAP	STANDARD:	PRT:	14 AA.
AC	044160;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	30-MAY-2000 (Rel. 39, Last annotation update)			
DE	30S RIBOSOMAL PROTEIN S19 (FRAGMENT).			
GN	PESS OR RPS19.			
OS	Prunus armeniaca phytoplasma.			
OC	Bacteriia: Firmicutes; Bacillus/Clostridium group; Molluscites;			
OC	Actinobacteria: Actinomycetes; Phytoplasma.			
OX	NCBI_TaxID=36589;			
RP	[1]			
RA	SEQUENCE FROM N.A.			
RX	MEDLINE=94350802; Pubmed=8071198;			
RA	Gundersen D.E., Lee I.M., Rehner S.A., Davis R.E., Kingsbury D.T.;			
RT	"Phylogeny of mycoplasma-like organisms (phytoplasmas): a basis for			
RT	their classification."			
RL	J. Bacteriol. 176:5244-5254(1994).			
CC	-1- FUNCTION: PROTEIN S19 FORMS A COMPLEX WITH S13 THAT BINDS STRONGLY			
CC	TO THE 16S RIBOSOMAL RNA (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC				
DR	EMBL; L26994; AAA83934.1;			
DR	InterPro; IPR002222; Ribosomal_S19.			
DR	PROSITE; PS00323; RIBOSOMAL_S19; PARTIAL.			
KW	Ribosomal protein; rRNA-binding.			
FT	NON_TER			
SQ	SEQUENCE 14 AA: 1756 MW: 4CC029EF8FFFE4A3 CRC64:			
Query Match				
Best Local Similarity 29.6% Score 24; DB 1; Length 14;				
Matches 2; Conservative 7; Mismatches 1; Indels 0; Gaps 0;				
OY	7 HDEQORELEE 16			
DB	4 HDKDKRKTK 13			
RESULT 2				
LE06_BIOGL	STANDARD:	PRT:	16 AA.	
AC	P80745;			
DT	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			

DE HEMOLYMPH 65 KDA LECTIN BG06 (FRAGMENT).  
 GN BG06.  
 OS Biomphalaria glabrata (Bloodfluke planorb).  
 CC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
 CC Planorbidae; Biomphalaria.  
 CX NCBI\_TaxID=6526;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=M-LINE; TISSUE=Hemolymph;  
 RX MEDLINE=97385165; PubMed=9238039;  
 RA Adema C.M., Hertel L.A., Miller R.D., Loker E.S.;  
 RT "A family of fibrinogen-related proteins that precipitates parasite-  
 RT derived molecules is produced by an invertebrate after infection."  
 CC Proc. Natl. Acad. Sci. U.S.A. 94:8691-8696(1997).  
 CC -1- FUNCTION: BINDS AND PRECIPITATES ANTIGENS OF THE PARASITE  
 CC ECHINOSTOMA PARASENSI.  
 CC -1- SUBCELLULAR LOCATION: SECRETED IN THE HEMOLYMPH.  
 CC -1- INDUCTION: BY INFECTION.  
 YN LECTIN.  
 YN NON\_TER 1 1  
 YN NON\_TER 16 16  
 YN SEQUENCE 16-AA; 1964 MW; A1665754589EF82C CRC64;  
 SQ  
 Query Match 29.6%; Score 24; DB 1; Length 16;  
 Best Local Similarity 44.4%; Pred. No. 5.2e+02;  
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 KRHEDOORE 13  
 1 1 1 1 1  
 Db 7 KDNDDOOND 15

RESULT 3  
 YNTP\_XANCP STANDARD; PRT; 14 AA.  
 ID YNTP\_XANCP  
 AC 003397;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1993 (Rel. 27, Last annotation update)  
 DE HYPOTHETICAL PROTEIN IN FRUK 5' REGION (ORF1) (FRAGMENT).  
 GN XMP.  
 OS Xanthomonas campestris (pv. campestris).  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xanthomonas.  
 CX NCBI\_TaxID=340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92011547; PubMed=1655739;  
 RA de Crecy-Lagard V., Bouvet O.M., Lejeune P., Danchin A.;  
 RT "Fructose catabolism in Xanthomonas campestris pv. campestris.  
 RT Sequence of the PTS operon, characterization of the fructose-specific  
 RT enzymes."  
 CC J. Biol. Chem. 266:18154-18161(1991).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M69242; AAA27600.1;  
 DR PIR: C40944; C40944.  
 DR Hypothetical protein.  
 KW NON\_TER 1 1  
 KW SEQUENCE 14 AA; 1585 MW; 72EA9AA3CE16CCDB CRC64;  
 SO

Query Match 25.9%; Score 21; DB 1; Length 14;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+03;  
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 9 EQORELE 15  
 1 1 1 1 1  
 Db 1 EQVRALE 7

RESULT 4  
 UNOL\_PINPS STANDARD; PRT; 15 AA.  
 ID UNOL\_PINPS  
 AC P81106;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 38, Last annotation update)  
 DE UNKNOWN PROTEIN FROM 2D-PAGE OF NEEDLES (S1247/S1248) (M150/M151)  
 DE (FRAGMENT).  
 OS Pinus pinaster (Maritime pine).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Coniferales; Pinaceae; Pinus.  
 CX NCBI\_TaxID=71647;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Needle;  
 RA Plomion C., Costa P., Bahrmann N., Frigerio J.M.;  
 RT "Genetic analysis of needle proteins in maritime pine. I. Mapping  
 RT dominant and codominant protein markers assayed on diploid tissue, in  
 RT a haploid-based genetic map."  
 CC Silvae Genetica 46:161-165(1997).  
 RN [2]  
 RP SEQUENCE.  
 RC TISSUE=Needle;  
 RX MEDLINE=99274088; PubMed=10344291;  
 RA Costa P., Plomion C., Bauw G., Dudos C., Bahrmann N., Kremer A.,  
 RA Frigerio J.M., Plomion C.;  
 RT "Separation and characterization of needle and xylem maritime pine  
 RT proteins."  
 CC Electrophoresis 20:1098-1108(1999).  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 5.5, ITS MW IS: 62 KDA.  
 FT NON\_TER 1 1  
 FT NON\_TER 15 15  
 FT SEQUENCE 15 AA; 1670 MW; 29CE44CD51E9FCF CRC64;  
 SO

OY 7 HDEQ 10  
 1 1 1 1 1  
 Db 1 HDEQ 4

RESULT 5  
 NEUX\_HUMAN STANDARD; PRT; 9 AA.  
 ID NEUX\_HUMAN  
 AC P04277;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 01-JUL-1989 (Rel. 11, Last annotation update)  
 DE NEUROTENSIN-RELATED PEPTIDE (NRP) (KINRENSIN).  
 OS Homo sapiens (Human), Bos taurus (Bovine), and  
 OS Oryctolagus cuniculus (Rabbit).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.  
 CX NCBI\_TaxID=9606, 9913, 9986;  
 RN [1]  
 RP SEQUENCE.  
 RC SPECIES=Human;  
 RX MEDLINE=86242180; PubMed=3087352;  
 RA Mogard M.H., Kobayashi R., Chen C.F., Lee T.D.,  
 RA Shively J.E., Walsh J.H.;  
 RT "The amino acid sequence of kinetensin, a novel peptide isolated from  
 RT peptin-treated human plasma: homology with human serum albumin,

Query Match 24.7%; Score 20; DB 1; Length 15;  
 Best Local Similarity 75.0%; Pred. No. 1.9e+03;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:13:12 ; Search time 12.75 Seconds  
(without alignments)  
95.592 Million cell updates/sec

Title: US-09-822-110-3

Perfect score: 81  
Sequence: 1 EIKAKRHDEQRELEE 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

al number of hits satisfying chosen parameters: 2758

Minimum DB seq length: 0  
Maximum DB seq length: 16

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR68:  
2: PIR1:  
3: PIR3:  
4: PIR4:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	30.9	7	2	alpha-myosin heavy
2	25	30.9	14	2	synaptosomal-assoc
3	22	27.2	12	2	T-cell antigen rec
4	22	27.2	15	2	beta-tubulin germ-
5	21	25.9	14	2	hypothetical prote
6	21	25.9	16	2	caldesmon - turkey
7	21	25.9	16	2	light-harvesting p
8	20	24.7	11	2	hypothetical 1.5K
9	20	24.7	12	2	beta-conglycinin a
10	20	24.7	14	2	hemocyanin chain 3
11	20	24.7	14	2	tubulin beta-3 cha
12	20	24.7	14	2	peptidylglycine mo
13	19	23.5	7	2	serum albumin - do
14	19	23.5	8	2	chlorophyll a/b-b1
15	19	23.5	9	2	chemical-sense-rel
16	19	23.5	11	2	glucocorticoidase
17	19	23.5	11	2	glucocorticoidase
18	19	23.5	12	2	hypothetical prote
19	19	23.5	12	2	lactose phosphora
20	19	23.5	14	2	gene C protein - E
21	19	23.5	14	2	hypothetical prote
22	19	23.5	15	2	urogenital tumor m
23	19	23.5	15	2	T-cell receptor be
24	19	23.5	15	2	T-cell receptor de
25	18	22.2	10	2	cytochrome-c oxida
26	18	22.2	13	2	enkephalin precurs
27	18	22.2	14	2	Ig kappa chain J s
28	18	22.2	15	2	hemoglobin beta ch
29	18	22.2	15	2	prostaglandin D-sy

30	18	22.2	15	2	PA0059	protein QF20021 -
31	18	22.2	15	2	S57584	T cell receptor V-
32	18	22.2	15	2	A53594	calnexin - mouse
33	18	22.2	16	2	C58503	proteoglycan assoc
34	18	22.2	16	2	S28213	glutathione transf
35	17	21.0	12	2	A35585	cytokinin-binding
36	17	21.0	12	2	A34858	proteinase E - bla
37	17	21.0	12	2	B47171	chondroitin sulfat
38	17	21.0	12	4	PC2122	aminotransferase c
39	17	21.0	13	2	J02309	hypothetical 1.6K
40	17	21.0	13	2	J02319	hypothetical 1.6K
41	17	21.0	13	2	PC1149	equinatoxin 1A - s
42	17	21.0	13	2	B47415	mannose-1-phosphat
43	17	21.0	14	2	A48389	leutinotoxin III -
44	17	21.0	14	2	PC7079	unidentified 27.2K
45	17	21.0	14	2	S65392	cytochrome-c oxida

## ALIGNMENTS

RESULT 1  
I46868  
alpha-myosin heavy chain : rabbit (fragment)  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 14-Feb-1997 #sequence\_revision 14-Feb-1997 #text\_change 05-Nov-1999  
C:Accession: I46868  
R:Friedman, D.J.; Umeda, P.K.; Sluhs, A.M.; Hsu, H.  
Proc. Natl. Acad. Sci. U.S.A. 81, 3044-3048, 1984  
A:Title: Characterization of genomic clones specifying rabbit alpha- and beta-ventric  
A:Accession: I46868  
A:Reference number: I46868; MUID:84221901  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-7 <FR1>  
A:Cross-references: GB:R01698; NID:q165538; PIDN:AAA31415.1; PID:q165539

Query Match  
Best Local Similarity 66.7%; Pred. No. 2.2e+05;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 KRHDEQ 10  
DB 2 KMHDEE 7

## RESULT 2

C44823  
synaptosomal-associated protein SNAP-25 peptide 8 - rabbit (fragment)

N:Alternate names: superprotein peptide 8

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 31-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 15-Jun-1996

C:Accession: C44823

R:Loewy, A.; Liu, W.S.; Baillinger, C.; Willard, M.B.

J. Neurosci. 11, 3412-3421, 1991

A:Title: The major 35S-methionine-labeled rapidly transported protein (superprotein)

A:Reference number: A44823; MUID:92044785

A:Accession: C44823

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-14 <LOB>

A:Experimental source: Visual tissue

A:Note: sequence extracted from NCBI backbone (NCBIP:64253)

C:Keywords: membrane trafficking

Query Match  
Best Local Similarity 62.5%; Pred. No. 8.1e+02;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 8 DEQRELE 15  
DB 11 : 11

Db 6 DEORELEE 13

## RESULT 3

S47360

T-cell antigen receptor VJ junction beta chain - human

C:Species: Homo sapiens (man)

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 05-Nov-1999

C:Accession: S47360

R:Lehner, P.J

Submitted to the EMBL Data Library, August 1994

A:Description: Human HLA-A0201 restricted recognition of influenza A is dominated by T

A:Reference number: S47355

A:Accession: S47360

A:Molecule type: mRNA

A:Status: preliminary

A:Residues: 1-12 &lt;LEH&gt;

A:Cross-references: EMBL:Z35684; NID:9527457; PIDN:CAM84753.1; PID:9527458

C:Keywords: T-cell receptor

## Query Match

27.2%; Score 22; DB 2; Length 12;

Best Local Similarity 57.1%; Pred. No. 1.9e+03;

Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 10 DEORELEE 16

Db 2 DEORELEE 8

## RESULT 4

A43839

beta-tubulin germ-cell isotype - African clawed frog (fragment)

C:Species: Xenopus laevis (African clawed frog)

C:Date: 10-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 11-Apr-1995

C:Accession: A43839

R:Bleker, J.J.; Yazdani-Bulky, M.

Differentiation 50, 15-23, 1992

A:Title: The multiple beta-tubulin genes of Xenopus: isolation and developmental express

A:Reference number: A43839; MUID:92347627

A:Accession: A43839

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-15 &lt;BLE&gt;

A:Experimental source: oocyte

A:Note: sequence extracted from NCBI backbone (NCBIP:110252)

## Query Match

27.2%; Score 22; DB 2; Length 15;

Best Local Similarity 50.0%; Pred. No. 2.4e+03;

Matches 4; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 9 DEORELEE 16

Db 1 DEORELEE 8

## RESULT 5

C40944

hypothetical protein (1-phosphofructokinase 5' region) - Xanthomonas campestris pv. camp

C:Species: Xanthomonas campestris pv. campestris

C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 18-Jun-1999

C:Accession: C40944

R:de Crecy-Lagard, V.; Bouvet, O.M.M.; Lejeune, P.; Danchin, A.

J. Biol. Chem. 266, 18154-18161, 1991

A:Title: Fructose catabolism in Xanthomonas campestris pv. campestris. Sequence of the F

A:Reference number: A40944; MUID:92011547

A:Accession: C40944

A:Molecule type: DNA

A:Residues: 1-14 &lt;DR3&gt;

A:Cross-references: GB:M69242; NID:9155366; PIDN:AAA27600.1; PID:9155367

C:Superfamily: fructose phosphotransferase multiphosphoryltransfer protein: phosphotrans

phosphotransferase system phosphotransferase-containing protein homology

Query Match 25.9%; Score 21; DB 2; Length 14;  
Best Local Similarity 71.4%; Pred. No. 3.1e+03;  
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 9 DEORELEE 15

Db 1 DEORELEE 7

## RESULT 6

A35552

caldesmon - turkey (fragment)

C:Species: Meleagris gallopavo (common turkey)

C:Date: 21-Sep-1990 #sequence\_revision 21-Sep-1990 #text\_change 24-Jun-1993

C:Accession: A35552

R:Velaz, L.; Ingraham, R.H.; Chalovich, J.M.

J. Biol. Chem. 265, 2929-2934, 1990

A:Title: Dissociation of the effect of caldesmon on the ATPase activity and on the bl

A:Reference number: A35552; MUID:90153926

A:Accession: A35552

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-16 &lt;VEL&gt;

## Query Match

25.9%; Score 21; DB 2; Length 16;

Best Local Similarity 66.7%; Pred. No. 3.6e+03;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 9 DEORELEE 14

Db 4 DEORELEE 9

## RESULT 7

F54226

light-harvesting protein B-830 beta-2 chain - Chromatium purpuratum (fragment)

C:Species: Chromatium purpuratum

C:Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 05-Jan-1996

C:Accession: F54226

R:Kerfeld, C.A.; Yeates, T.O.; Thorndyke, J.P.

Biochemistry 33, 2178-2184, 1994

A:Title: Purification and characterization of the peripheral antenna of the purple-su

A:Reference number: A54226; MUID:94162224

A:Accession: F54226

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-16 &lt;KER&gt;

C:Keywords: antenna complex; light-harvesting polypeptide

## Query Match

25.9%; Score 21; DB 2; Length 16;

Best Local Similarity 44.4%; Pred. No. 3.6e+03;

Matches 4; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 8 DEORELEE 16

Db 8 DEORELEE 16

## RESULT 8

JQ2307

hypothetical 1.5K protein - tomato chloroplast (strain Toko)

C:Species: Chloroplast Lycopersicon esculentum (tomato)

C:Date: 10-Mar-1994 #sequence\_revision 28-Oct-1994 #text\_change 23-Mar-1995

C:Accession: JQ2307

R:Kawagoe, Y.; Kikuta, Y.

Theor. Appl. Genet. 81, 13-20, 1991

A:Title: Chloroplast DNA evolution in potato (Solanum tuberosum L.).

A:Reference number: JQ2306

A:Accession: JQ2307

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:17:22 ; Search time 12.54 Seconds  
(without alignments)  
89.726 Million cell updates/sec

Title: US-09-822-110-5

Perfect score: 253  
Sequence: 1 PPKAKKPVSLIRDLTEAM.....QRELEEEENSDELDHSHT 50

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

total number of hits satisfying chosen parameters: 149863

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents: MA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
%score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50.5	20.0	30	2	US-08-619-198-7
2	50	19.8	35	1	US-08-182-175A-53
3	50	19.8	35	1	US-08-474-633A-62
4	50	19.8	35	5	PCT-US92-06412-53
5	46	18.2	34	2	US-08-690-011A-53
6	46	18.2	35	4	US-09-082-279B-72
7	45	17.8	28	1	US-08-182-175A-6
8	45	17.8	28	5	PCT-US92-06412-6
9	45	17.8	35	4	US-09-248-588-26
10	44	17.4	20	3	US-08-199-776-21
11	44	17.4	20	3	US-08-663-731-21
12	44	17.4	20	3	US-08-879-338-21
13	44	17.4	20	5	PCT-US95-02044-21
14	44	17.4	21	1	US-08-199-776-5
15	44	17.4	21	3	US-08-663-731-5
16	44	17.4	21	3	US-08-879-338-5
17	44	17.4	21	3	US-08-879-338-26
18	44	17.4	21	5	PCT-US95-02044-5
19	44	17.4	28	4	US-09-082-279B-74
20	44	17.4	43	4	US-09-015-030-7
21	43.5	17.2	26	2	US-08-690-011A-3
22	43	17.0	28	1	US-08-182-175A-49
23	43	17.0	28	1	US-08-474-633A-58
24	43	17.0	28	5	PCT-US92-06412-49
25	43	17.0	35	1	US-08-182-175A-41
26	43	17.0	35	1	US-08-474-633A-38
27	43	17.0	35	5	PCT-US92-06412-41

28	43	17.0	36	4	US-09-082-279B-860	Sequence 860, App
29	43	17.0	33	2	US-08-499-676A-29	Sequence 29, App1
30	42.5	16.8	34	2	US-08-690-011A-57	Sequence 57, App1
31	42.5	16.8	49	4	US-08-866-928B-10	Sequence 10, App1
32	42.5	16.8	49	4	US-08-866-928B-11	Sequence 11, App1
33	42.5	16.8	49	4	US-08-866-928B-12	Sequence 12, App1
34	42	16.6	26	2	US-08-523-125-2	Sequence 2, App1
35	42	16.6	26	3	US-08-660-561A-2	Sequence 2, App1
36	42	16.6	28	1	US-08-182-175A-4	Sequence 4, App1
37	42	16.6	28	1	US-08-056-200-13	Sequence 13, App1
38	42	16.6	28	2	US-08-800-644-13	Sequence 13, App1
39	42	16.6	28	5	PCT-US92-06412-4	Sequence 4, App1
40	42	16.6	45	3	US-08-816-346-12	Sequence 12, App1
41	42	16.6	45	4	US-09-335-411-12	Sequence 12, App1
42	41.5	16.4	33	2	US-08-690-011A-38	Sequence 38, App1
43	41	16.2	13	3	US-08-199-776-25	Sequence 25, App1
44	41	16.2	13	3	US-08-663-731-25	Sequence 25, App1
45	41	16.2	13	3	US-08-879-338-25	Sequence 25, App1

#### ALIGNMENTS

RESULT 1  
US-08-619-198-7  
Sequence 7, Application US/08619198  
Patent No. 5885831  
GENERAL INFORMATION:  
APPLICANT: Young, Michael W.  
APPLICANT: Senegal, Amila  
APPLICANT: Vosschall, Leslie B.  
APPLICANT: Price, Jeffrey L.  
APPLICANT: Myers, Michael  
TITLE OF INVENTION: NUCLEAR LOCALIZATION FACTOR ASSOCIATED  
TITLE OF INVENTION: WITH CIRCADIAN RHYTHMS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/619,198  
FILING DATE: 20-MAR-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-128A CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEetical: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Drosophila melanogaster  
US-08-619-198-7

Query Match 20.0%; Score 50.5; DB 2; Length 30;  
Best Local Similarity 47.6%; Pred. No. 6.6;  
Matches 10; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

OY 28 DEQORELEEE-DENDDELD 47  
DB 7 DQRRHRLNHEGEDEDEVE 27

RESULT 2  
US-08-182-175A-53  
Sequence 53, Application US/08182175A  
Patent No. 5559223

GENERAL INFORMATION:  
APPLICANT: Saverio Carl Falco  
APPLICANT: Sharon J. Keeler  
APPLICANT: Janet A. Rice  
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containing F  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. du Pont de Nemours and Company  
STREET: 1007 Market Street  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: USA  
ZIP: 19898

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh System, 6.0  
SOFTWARE: Microsoft Word, 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/182,175A  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/743,006  
FILING DATE: 9 August 1991

ATTORNEY/AGENT INFORMATION:  
NAME: Linda Axamechy Floyd  
REGISTRATION NUMBER: 33,692  
REFERENCE/DOCKET NUMBER: BB-1031  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (302) 992-4929  
TELEFAX: (302) 892-7949  
TELEX: 835420

INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-182-175A-53

Query Match 19.8%; Score 50; DB 1; Length 35;  
Best Local Similarity 34.6%; Pred. No. 9;  
Matches 9; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

OY 18 EAMEIKAKRHDEQORELEEEENSDE 43  
DB 6 KAMEEKMKMEKMKMEKMKMEK 31

RESULT 3

US-08-474-633A-62  
Sequence 62, Application US/08474633A  
Patent No. 5773691  
GENERAL INFORMATION:  
APPLICANT: E. I. du Pont de Nemours and  
APPLICANT: COMPANY  
TITLE OF INVENTION: CHIMERIC GENES AND

TITLE OF INVENTION: METHODS FOR INCREASING  
TITLE OF INVENTION: INCREASING THE LYSINE  
TITLE OF INVENTION: AND THREONINE CONTENT  
TITLE OF INVENTION: OF THE SEEDS OF PLANTS  
NUMBER OF SEQUENCES: 107  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. du Pont de Nemours  
STREET: 1007 MARKET STREET  
CITY: WILMINGTON  
STATE: DELAWARE  
COUNTRY: U.S.A.  
ZIP: 19898

COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MICROSOFT WORD VERSION 2.0C  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,633A  
FILING DATE:  
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:  
NAME: BARBARA C. SIEGELL  
REGISTRATION NUMBER: 30,684  
REFERENCE/DOCKET NUMBER: BB-1037-C  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 302-992-4931  
TELEFAX: 302-773-0164  
TELEX: 835420

INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 35 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-474-633A-62

Query Match 19.8%; Score 50; DB 1; Length 35;  
Best Local Similarity 34.6%; Pred. No. 9;  
Matches 9; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

OY 18 EAMEIKAKRHDEQORELEEEENSDE 43  
DB 6 KAMEEKMKMEKMKMEKMKMEK 31

RESULT 4  
PCT-US92-06412-53  
Sequence 53, Application PC/TUS9206412

GENERAL INFORMATION:  
APPLICANT: Saverio Carl Falco  
APPLICANT: Sharon J. Keeler  
APPLICANT: Janet A. Rice  
TITLE OF INVENTION: Synthetic Storage Proteins with Defined Structure Containin  
NUMBER OF SEQUENCES: 113  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: E. I. du Pont de Nemours and Company  
STREET: 1007 Market Street  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: USA  
ZIP: 19898

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh System, 6.0  
SOFTWARE: Microsoft Word, 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/06412  
FILING DATE: 19920807  
CLASSIFICATION: 530

Result	No.	Score	Query	Match	Length	DB	ID	Description
1	62	24.5	49	22	22	AA021894	Peptide #8328 encco	
2	62	24.5	49	22	22	AA038820	Peptide #12257 eno	
3	61	24.1	37	18	18	AA045849	Peptide conjugate	
4	56	22.1	41	18	18	AA006736	Calreticulin C-dom	
5	54	21.3	31	22	22	AA033848	Peptide #7885 encco	
6	54	21.3	41	22	22	AA016668	Peptide #3102 encco	
7	54	21.3	41	22	22	AA029152	Peptide #3189 encco	
8	54	21.3	41	22	22	AA004384	Peptide #3066 encco	
9	54	21.3	50	22	22	AA015044	Peptide #1478 encco	
10	54	21.3	50	22	22	AA027486	Peptide #1523 encco	
11	54	21.3	50	22	22	AA002775	Peptide #1457 encco	

	12	50.5	20.0	30	20	AAY00885	Timeless gene protein
XX	13	50	19.8	35	16	AAK78253	SSP 5.10.10.5.
XX	14	50	19.8	35	19	AAAM60519	Synthetic storage
XX	15	50	19.8	35	20	AAAB87751	Synthetic lysine-T
XX	16	50	19.8	36	14	AAAK32696	SSP polypeptide pr
XX	17	50	19.8	39	22	AAAM20640	Peptide #7074 enco
XX	18	50	19.8	39	22	AAAM35667	Peptide #9704 enco
XX	19	50	19.8	41	22	AAAM15885	Peptide #1719 enco
XX	20	49	19.4	24	21	AAAB26895	Proteasome alpha7
XX	21	49	19.4	24	22	AAAB97284	Murine 20S proteas
XX	22	49	19.4	29	22	AAAM33060	Peptide #7097 enco
XX	23	48.5	19.2	41	22	AAAM18959	Peptide #5393 enco
XX	24	48.5	19.2	41	22	AAAM31544	Peptide #5581 enco
XX	25	48	19.0	33	16	AAAR87260	KEKE-motif of huma
XX	26	47	18.6	23	16	AAAM21451	HSV 2 glycoprotein
XX	27	47	18.6	34	19	AAAM63041	Part of the protea
XX	28	47	18.6	38	22	AAAM32445	Peptide #6282 enco
XX	29	47	18.6	42	16	AAAR87208	Heat shock protein
XX	30	47	18.6	43	21	AAAB34541	Human secreted prot
XX	31	47	18.6	45	22	AAAM27749	Peptide #1786 enco
XX	32	47	18.6	45	22	AAAM03029	Peptide #1711 enco
XX	33	46	18.2	28	14	AAAR31234	Prepro-Chytotropin
XX	34	46	18.2	35	21	AAAB87821	Core polypeptide f
XX	35	46	18.2	35	22	AAAB77076	Core polypeptide T
XX	36	46	18.2	42	21	AAAG47591	Arabidopsis thaliaa
XX	37	46	18.2	42	21	AAAG47590	Arabidopsis thaliaa
XX	38	46	18.2	49	21	AAAG47589	Arabidopsis thaliaa
XX	39	45	17.8	28	14	AAAR31984	SSP10 polypeptide,
XX	40	45	17.8	35	20	AAAV29676	S. pneumoniae anti
XX	41	45	17.8	47	22	AAAM17064	Peptide #3498 enco
XX	42	45	17.8	47	22	AAAM29558	Peptide #3395 enco
XX	43	45	17.8	47	22	AAAM04766	Peptide #3448 enco
XX	44	44	17.4	20	16	AAAR82651	HML-1 alpha-E chain
XX	45	44	17.4	21	16	AAAR82636	HML-1 alpha-E chain

ALIGNMENTS

RESULT	1						
ID	AAAM21894						
AC	AAAM21894 standard; protein; 49 AA.						
XX	AAAM21894;						
DT	12-OCT-2001 (first entry)						
XX	Peptide #8328 encoded by probe for measuring cervical gene expression.						
XX	Probe; human; microarray; gene expression; cervical epithelial cell;						
KW	cervical cancer.						
OS	Homo sapiens.						
XX	WO200157278-A2.						
PN							
PD	09-AUG-2001.						
PF	30-JAN-2001; 2001WO-US00670.						
XX							
PR	04-FEB-2000; 2000US-0180312.						
PR	26-MAY-2000; 2000US-0207456.						
PR	30-JUN-2000; 2000US-0608408.						
PR	03-AUG-2000; 2000US-0632366.						
PR	21-SEP-2000; 2000US-0234687.						
PR	27-SEP-2000; 2000US-0236359.						
PR	04-OCT-2000; 2000GB-0024263.						
XX							
PA	(MOLE-) MOLECULAR DYNAMICS INC.						
PI	Penn SG, Hanzel DK						
XX	Chen W, Rank DR;						
DR	WPI: 2001-488901/53.						

XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells -  
 XX  
 PS Claim 27; SEQ ID No 26720; 487pp; English.  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENPs; see A110068-A128459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [ftp.wipo.int/pub/published\\_pcl\\_sequences](http://wipo.int/pub/published_pcl_sequences).  
 CC  
 SO Sequence 49 AA;  
 Query Match 24.5%; Score 62; DB 22; Length 49;  
 Best Local Similarity 46.7%; Pred. No. 1.3;  
 Matches 14; Conservative 3; Mismatches 13; Indels 0; Gaps 0;  
 17 TEAMEIKAKRHDEQORELEEEENSDEDEL 46  
 16 lvakvakaekpeeeesaeeesndedev 45  
 RESULT 2  
 AAM38220  
 ID AAM38220 standard; Protein; 49 AA.  
 AC AAM38220;  
 XX 17-OCT-2001 (first entry)  
 DE Peptide #12257 encoded by probe for measuring placental gene expression.  
 XX  
 KW Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder;  
 OS Homo sapiens.  
 PN WC200157272-A2.  
 XX 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US00663.  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 WPI; 2001-488897/53.  
 DR WPI; 2001-488897/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 XX  
 PS Claim 27; SEQ ID No 38489; 654pp; English.  
 CC The present invention relates to single exon nucleic acid probes (SENPs;  
 CC see A131315-A157546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 CC  
 SO Sequence 49 AA;  
 Query Match 24.5%; Score 62; DB 22; Length 49;  
 Best Local Similarity 46.7%; Pred. No. 1.3;  
 Matches 14; Conservative 3; Mismatches 13; Indels 0; Gaps 0;  
 17 TEAMEIKAKRHDEQORELEEEENSDEDEL 46  
 16 lvakvakaekpeeeesaeeesndedev 45  
 RESULT 3  
 AAM45249  
 ID AAM45249 standard; peptide; 37 AA.  
 AC AAM45249;  
 XX 06-MAY-1998 (first entry)  
 DE Peptide conjugate for transporting substances across cell membranes.  
 XX  
 KW Hydrophilic; lipophilic; lipid bilayer; peptide conjugate; transporting;  
 KW pH-dependent transition; cell membrane.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 1 /label= bala  
 FT /note= "beta-Alanine"  
 FT Modified-site 37 /note= "conjugated to a Morpholino oligo"  
 XX  
 PN WC9740854-A2.  
 PD 06-NOV-1997.  
 XX 30-APR-1997; 97WO-US07335.  
 PF 23-OCT-1996; 96US-0028609.  
 PR 01-MAY-1996; 96US-0016347.  
 XX  
 PA (ANT-) ANTIVIRALS INC.  
 PI Summerton JE, Weller DD;  
 WPI; 1997-549493/50.  
 DR WPI; 1997-549493/50.  
 XX Polypeptide composition able to reversibly transit between  
 PT lipophilic and hydrophilic forms - is useful for transporting  
 PT compounds across lipid layers  
 XX  
 PS Example 8; Figure 13C; 73pp; English.  
 XX  
 CC This sequence represents a peptide linked to a Morpholino antisense  
 CC oligo conjugate. The invention relates to new peptide conjugates  
 CC which are useful for transporting a compound from a low pH environment  
 CC across a lipid layer to a higher pH aqueous compartment and comprise:  
 CC (a) a polypeptide (PP) containing one or more pairs of carboxyl groups;  
 CC where: (i) the carboxyl groups of a pair are separated by 0, 2 or 3  
 CC amino acids; (ii) the PP has a length of 8-100 amino acid residues;  
 CC (iii) the PP can undergo a reversible transition between a lipophilic  
 CC form (which can partition from the low pH environment into the lipid  
 CC layer) and a hydrophilic form (which can partition preferentially from  
 CC the lipid layer into the higher pH compartment); and (iv) the PP contains  
 CC an initiator moiety at one end region, to facilitate entry of the end  
 CC region into the lipid layer; where the PP is able to traverse the lipid  
 CC layer from the low-pH to the higher-pH compartment; and (b) the compound



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:18:12 ; Search time 22.07 Seconds  
(without alignments)  
331.383 Million cell updates/sec

Title: US-09-822-110-5  
Perfect score: 253  
Sequence: 1 PFIRNAKPVSIIRLDITTEAM.....ORELEEEENSDDELDHSHT 50

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
Total number of hits satisfying chosen parameters: 27079

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mmc:\*  
9: sp\_organella:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	22.1	41	5	09TX57
2	52	20.6	45	5	09VFU5
3	49.5	19.6	50	11	096219
4	49	19.4	41	10	096414
5	49	19.4	44	12	069202
6	48.5	19.2	44	12	069199
7	48	19.0	44	10	003996
8	48	18.8	44	5	09VR85
9	47.5	18.8	44	4	016264
10	47	18.6	44	12	069203
11	47	18.6	44	12	069192
12	47	18.6	44	12	069198
13	47	18.6	46	5	009963
14	46.5	18.4	50	10	09C6E9
15	46	18.2	44	5	09N7M4
16	44	17.4	44	12	069195
17	43	17.0	37	11	062620
18	43	17.0	48	4	012978
19	43	17.0	50	2	006970

20	43	17.0	50	12	09JUE6	09JUE6 vaccinia vi
21	42	16.6	43	10	09MSU2	09MSU2 nicotiana t
22	42	16.6	44	12	0993M8	0993M8 human immu
23	42	16.6	45	2	09KED2	09KED2 bacillus ha
24	42	16.6	46	5	020835	020835 caenorhabdi
25	41.5	16.4	45	10	09MSU1	09MSU1 nicotiana t
26	41.5	16.4	50	2	099725	099725 staphylococ
27	41	16.2	44	5	09V7I6	09V7I6 drosophila
28	41	16.2	49	13	090340	090340 cyprinus ca
29	41	16.2	49	13	090340	090340 cyprinus ca
30	40.5	16.0	35	4	09UE13	09UE13 homo sapien
31	40.5	16.0	43	13	090347	090347 colutrix co
32	40.5	16.0	50	10	09LUZ2	09LUZ2 arabidopsis
33	40	15.8	47	12	09MX4	09MX4 human paral
34	40	15.8	49	6	09BDE6	09BDE6 bos taurus
35	40	15.8	50	12	064879	064879 unidentified
36	40	15.8	50	12	09MX3	09MX3 human paral
37	39	15.4	16	11	090UW8	090UW8 cavia (guin
38	39	15.4	20	4	015969	015969 homo sapien
39	39	15.4	44	11	060500	060500 cricetus
40	39	15.4	44	12	P87665	P87665 duck adenov
41	39	15.4	48	5	09VWS6	09VWS6 drosophila
42	39	15.4	50	4	09P0K4	09P0K4 homo sapien
43	38	15.0	23	6	029394	029394 canis fami
44	38	15.0	34	2	069770	069770 pseudomonas
45	38	15.0	48	2	09ZL87	09ZL87 helicobacte

## ALIGNMENTS

RESULT 1  
ID 09TX57 PRELIMINARY: PRT: 41 AA.  
AC 09TX57:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE CADESMON HOMOLOG (FRAGMENT).  
OS Mytilus.  
OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytilloidea;  
OC Mytilloidea; Mytilidae.  
OX NCBI\_TaxID=6548;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94265000; PubMed=8205392;  
RA Cizmadi A.M., Bonet-Kerrache A., Nyitray L., Mornet D.;  
RT "Purification and properties of caldesmon-like protein from molluscan  
RT smooth muscle."  
RL Comp. Biochem. Physiol. 108B:59-63(1994).  
FT NON\_TER 1 1  
FT NON\_TER 41 41  
SO SEQUENCE 41 AA; 5512 MW; DD511D27BA04A555 CRC64;

Query Match 22.1%; Score 56; DB 5; Length 41;  
Best Local Similarity 41.7%; Pred. No. 47;  
Matches 10; Conservative 7; Mismatches 7; Indels 0; Gaps 0;  
OY 21 EIKAKRHDEQORELEEEENSD 44  
DB 18 EDRRRERDERRERREERKRKE 41  
RESULT 2  
ID 09VFU5 PRELIMINARY: PRT: 45 AA.  
AC 09VFU5:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE CG14371 PROTEIN.

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga C., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu I., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrel J., Fleischmann W.,  
 RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glisler A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jatala M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reiner K., Remington K., Sanders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RL EMBL: AE003700; AAF54954.1;  
 RL FlyBase: FBgn0040555; CG14371.  
 SO SEQUENCE 45 AA; 5235 MW; 93A8CBA8E0F54082 CRC64;

Query Match 20.6%; Score 52; DB 5; Length 45;  
 Best Local Similarity 27.8%; Pred. No. 1.3e+02;  
 Matches 10; Conservative 11; Mismatches 15; Indels 0; Gaps 0;

OY 10 SLINDLITAMEIKAKRHDEQORLEEFEEENSDEDE 45  
 ID 062619 PRELIMINARY; PRT; 50 AA.  
 AC 062619;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE FETAL TROPONIN T 3 (FRAGMENT).  
 OC Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE, SKELETAL;  
 RX MEDLINE=93345743; PubMed=83444466;  
 RA Briggs M.M., Schachar F.,  
 RT "Origin of fetal tropoin T: developmentally regulated splicing of a  
 RT new exon in the fast tropoin T gene."  
 RL Dev. Biol. 158:503-509(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE, SKELETAL;  
 RA Briggs M.M.;  
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U04980; AAA16033.1;  
 KW Alternative splicing  
 FT NON\_TER 50  
 SO SEQUENCE 50 AA; 5982 MW; A4ACC3068914925D CRC64;

Query Match 19.4%; Score 49; DB 10; Length 50;  
 Best Local Similarity 36.1%; Pred. No. 2.5e+02;  
 Matches 13; Conservative 5; Mismatches 11; Indels 7; Gaps 1;

OY 17 TEAMEIKAKRHDEQORELE-----EEENSDEDE 45  
 ID 096414 PRELIMINARY; PRT; 41 AA.  
 AC 096414;  
 DT 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE BETA-TUBULIN 3 (FRAGMENT).  
 GN TUBB.  
 OS Daucus carota (Carrot).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.  
 OC NCBI\_TaxID=4039;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, KINTOKI;  
 RA Okamura S., Sonehara K., Naito K., Ohkawa H., Kuramori S., Tatsuta M.,  
 RA Mimazono M.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U64430; AAB47936.1;  
 DR Mendel: 13640; Dauga; TUBB; 13640.  
 DR InterPro: IPR003008; Tubulin\_FtsZ.  
 DR Pfam: PF00091; tubulin; 1.  
 KW GTP-binding.  
 FT NON\_TER 1  
 SO SEQUENCE 41 AA; 4945 MW; A5236AAE45F10274 CRC64;

Query Match 19.4%; Score 49; DB 10; Length 41;  
 Best Local Similarity 32.4%; Pred. No. 2.4e+02;  
 Matches 11; Conservative 7; Mismatches 16; Indels 0; Gaps 0;

OY 10 SLINDLITAMEIKAKRHDEQORLEEFEEENSDEDE 43  
 ID 069202 PRELIMINARY; PRT; 44 AA.  
 AC 069202;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:19:17 ; Search time 10.1 Seconds

(without alignments)  
181.509 Million cell updates/sec

Title: US-09-822-110-5

Sequence: 1 PF1KAKRVSILRDILITEAM.....QRELEEEENDEDELSHT 50

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

total number of hits satisfying chosen parameters: 3562

Minimum DB seq length: 0

Maximum DB seq length: 50

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	15.4	46	1 DIUH_PERAM	P41538 periplaneta
2	39	15.4	46	1 DIUH_ZOONE	P82707 zoothermopsi
3	38	15.0	39	1 PS11_PSEFL	P80694 pseudomonas
4	38	15.0	44	1 DIUH_MUSDO	P41537 musca domes
5	36	14.2	49	1 OSTC_XENLA	P40147 xenopus lae
6	36	14.2	49	1 OSTC_XENLA	P41593 spiroplasma
7	35	13.8	45	1 RK34_CVACA	O19912 cyanidium c
8	35	13.8	48	1 OSTC_DRONO	P15504 dromaeus no
9	35	13.8	50	1 OM05_YEAST	P80967 saccharomyc
10	34.5	13.6	46	1 DIUH_DIPPU	P82373 diptoptera
11	34	13.4	23	1 VG22_BPT2	P21596 bacterioph
12	34	13.4	23	1 VG22_BPT2	P21597 bacterioph
13	33.5	13.2	26	1 RECX_YERPE	P37867 yerinia pe
14	33.5	13.2	46	1 DIUH_LOCOMI	P23465 locusta mig
15	33	13.0	13	1 PRC_SALTY	P33669 salmoneilla
16	33	13.0	40	1 SAVV_PHYSA	P01144 phyllomedus
17	33	13.0	42	1 YOAH_SALTY	P56505 salmoneilla
18	33	13.0	46	1 ALAT_MACEU	P38027 macropus eu
19	32.5	12.8	46	1 VIL_BPT3	P20833 bacterioph
20	32	12.6	40	1 YCO4_ARCFU	O29064 archaeglob
21	32	12.6	49	1 OSTC_FELCA	P02821 felis silve
22	32	12.6	50	1 SMD2_PIG	Q29329 sus scrofa
23	31	12.3	30	1 ACBI_DIGLA	P81624 digitalis 1
24	31	12.3	37	1 VPU_HV128	P08807 human immun
25	31	12.3	39	1 AKH2_SCHGR	P35808 schistocerc
26	31	12.3	44	1 BAB4_BABBO	P14201 babesia bov
27	31	12.3	49	1 OSTC_MACFA	P02809 macaca fasc
28	30.5	12.1	33	1 ACT_DICVI	Q24733 dictyocaulu
29	30.5	12.1	37	1 GL36_LUPAN	P09390 lupinus ang
30	30.5	12.1	37	1 RL36_MYCCE	P17420 mycoplasma
31	30.5	12.1	45	1 VPU_HV1N5	P08804 human immun
32	30	11.9	21	1 FTBB_CEREL	P14466 cervus elap
33	30	11.9	43	1 ILVY_SALTY	P05988 salmoneilla

34	30	11.9	47	1 PR4_PHAVU	O09020 phaseolus v
35	30	11.9	49	1 COX4_PARDE	P77921 paracoccus
36	30	11.9	50	1 X234_HAEIN	P43967 haemophilus
37	29.5	11.7	39	1 FUC3_RAT	P80349 rattus norv
38	29.5	11.7	40	1 RRPO_LSV	P27328 illy sympto
39	29	11.5	20	1 OXLA_AGRH	P81382 agkistrodon
40	29	11.5	20	1 YOAH_KLEAE	P56506 klebsiella
41	29	11.5	22	1 ODPX_BOVIN	P22439 bos taurus
42	29	11.5	37	1 TCTE_TRYBB	P35758 trypanosoma
43	29	11.5	47	1 V820_METJA	O58230 methanococc
44	29	11.5	48	1 CATB_COTJA	P81494 coturnix co
45	29	11.5	49	1 OSTC_CANFA	P81455 canis fam11

## ALIGNMENTS

RESULT	ID	DIUH_PERAM	STANDARD	PRT	46 AA.
1	DIUH_PERAM	DIUH_PERAM	STANDARD	PRT	46 AA.
AC	P41538:				
DT	01-NOV-1995 (Rel. 32, Created)				
DT	01-NOV-1995 (Rel. 32, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	DIURETIC HORMONE (DH) (DIURETIC PEPTIDE) (DP).				
OS	Periplaneta americana (American cockroach).				
OC	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:				
OC	Pterygota: Neoptera: Orthopteroidea: Dictyoptera: Blattaria:				
OC	Blattoidea: Blattidae: Periplaneta.				
OX	NCBI_Taxid=6978:				
RN	[1]				
RP	SEQUENCE.				
RX	MELINE=93174045; PubMed=1337794;				
RA	Kay L., Patel M., Coast G.M., Totty N.F., Mallet A.I.,				
RA	Goldsworthy G.J.,				
RT	"Isolation, characterization and biological activity of a CRF-related				
RT	diuretic peptide from Periplaneta americana L.";				
RL	Regul. Pept. 42:111-122(1992).				
CC	-1- FUNCTION: REGULATION OF FLUID SECRETION. STIMULATES PRIMARY URINE				
CC	SECRETION BY MALPIGHIAN TUBULES AND CAUSES A DOSE-DEPENDENT				
CC	STIMULATION OF CAMP LEVELS IN THE TUBULES.				
CC	-1- SIMILARITY: BELONGS TO THE SAUVAGINE/CORTICOTROPIN-RELEASING				
CC	FACTOR/UBOTENSIN I FAMILY OF PEPTIDES.				
DR	PIR: A48542; A48542.				
DR	InterPro: IPR000187; CRF.				
DR	Pfam: PF00473; CRF; 1.				
DR	SMART: SM00039; CRF; 1.				
DR	PROSITE: PS00511; CRF; 1.				
KW	Hormone; Amidation.				
FT	MOD_RES 46				
SO	SEQUENCE 46 AA; 5285 MW; 6938D47CB5AF2BA5 CRC64;				
Query Match					
Oy	1	PF1KAKRVSILRDILITEAMEIKAKRHDEQRELEEE	39	15.48; Score 39; DB 1; Length 46;	
Best Local Similarity, 23.1%; Pred. No. 5.6e+02;					
Matches 9; Conservative 10; Mismatches 18; Indels 2; Gaps 1;					
Db	5	PSLSTVPLDVLKRL-LLEIARRRMSODQIOANRE	41		
RESULT 2					
ID	DIUH_ZOONE	STANDARD	PRT	46 AA.	
AC	P82707:				
DT	20-AUG-2001 (Rel. 40, Created)				
DT	20-AUG-2001 (Rel. 40, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	DIURETIC HORMONE (DH) (DIURETIC PEPTIDE) (DP).				
OS	Zoothermopsis nevadensis (Dampwood termite).				
OC	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:				
OC	Pterygota: Neoptera: Orthopteroidea: Dictyoptera: Isoptera:				

OC Termopsidae; Zootermopsids.  
 OX NCBI\_TaxID=136037;  
 RN [1]  
 RP SEQUENCE, FUNCTION, MASS SPECTROMETRY, AND AMIDATION.  
 RC TISSUE-Head;  
 RX MEDLINE=21107389; PubMed=11179807;  
 RA Baldwin D.C., Schegg K.M., Furuya K., Lehmberg E., Schooley D.A.;  
 RT "Isolation and identification of a diuretic hormone from Zootermopsids nevalensis."  
 RL Peptides 22:147-152(2001).  
 CC -1- FUNCTION: REGULATION OF FLUID SECRETION. STIMULATES PRIMARY URINE SECRETION BY MALPIGHIAN TUBULES AND CAUSES A DOSE-DEPENDENT STIMULATION OF CAMP LEVELS IN THE TUBULES.  
 CC -1- MASS SPECTROMETRY: MW=5329.0; METHOD=MALDI.  
 CC -1- SIMILARITY: BELONGS TO THE SAUVAGINE/CORTICOTROPIN-RELEASING FACTOR/UROTENSIN I FAMILY OF PEPTIDES.  
 CC Interpro; IPR000187; CRF.  
 DR Interpro; IPR003621; Diuric\_hormn.  
 DR Pfam; PF00473; CRF; 1.  
 DR Prodom; PD014750; Diuric\_hormn; 1.  
 DR SMART; SM00039; CRF; 1.  
 DR PROSITE; PS00511; CRF; 1.  
 DR Hormone; Amidation.  
 KM MOD\_RES 46  
 FT MOD\_RES 46  
 SQ SEQUENCE 46 AA; 5329 MW; 693BC02E11F151A5 CRC64;

Query Match 15.4%; Score 39; DB 1; Length 46;  
 Best Local Similarity 23.1%; Pred. No. 5.6e+02;  
 Matches 9; Conservative 10; Mismatches 18; Indels 2; Gaps 1;  
 OY 1 PFKNKAPVSIPLDLITEAMEIKARHDEQRELEEE 39  
 DB 5 PSLSTVNPDLVLRQL--LLEIARRRMQSQIOANRE 41

RESULT 3  
 PS11\_PSEFL STANDARD; PRT; 39 AA.  
 AC P80694;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PHOSPHATE STARVATION INDUCIBLE PROTEIN 1 (PS11) (FRAGMENT).  
 OS Pseudomonas fluorescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=294;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=AG1;  
 RX MEDLINE=97237729; PubMed=9084184;  
 RA Leopold K., Jacobsen S., Nybroe O.;  
 RT "A phosphate-starvation-inducible outer-membrane protein of Pseudomonas fluorescens Ag1 as an immunological phosphate starvation marker."  
 RL Microbiology 143:1019-1027(1997).  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 KW Outer membrane.  
 FT NON\_TER 39  
 SQ SEQUENCE 39 AA; 4290 MW; 02280E11312E4ACF-CRC64;

Query Match 15.0%; Score 38; DB 1; Length 39;  
 Best Local Similarity 28.6%; Pred. No. 5.9e+02;  
 Matches 8; Conservative 9; Mismatches 11; Indels 0; Gaps 0;

OY 12 LRDLLTEAMEIKARHDEQRELEEE 39  
 DB 6 LIDMLKANGQISASQYTELOAEIAKQK 33

RESULT 4

DIUH\_MUSDO STANDARD; PRT; 44 AA.  
 AC P41537;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DE DIURETIC HORMONE (DH) (DIURETIC PEPTIDE) (DP).  
 OS Musca domestica (House fly), and Stomoxys calcitrans (Stable fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Muscoidea; Muscidae; Musca.  
 OX NCBI\_TaxID=7370, 35570;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=95083479; PubMed=7991460;  
 RA Clotiens F.L., Holman G.M., Coast G.M., Totty N.F., Hayes T.K., Kay I., Maillet A.I., Wright M.S., Chung J.-S., Truong O., Bull D.L.;  
 RT "Isolation and characterization of a diuretic peptide common to the house fly and stable fly."  
 RL Peptides 15:971-979(1994).  
 CC -1- FUNCTION: REGULATION OF FLUID SECRETION. STIMULATES PRIMARY URINE SECRETION BY MALPIGHIAN TUBULES AND CAUSES A DOSE-DEPENDENT STIMULATION OF CAMP LEVELS IN THE TUBULES. MAY ACT AS CLEARANCE PEPTIDE IN THAT IT MAY REMOVE METABOLIC WASTE FROM THE HEMOLYMPH.  
 CC -1- SIMILARITY: BELONGS TO THE SAUVAGINE/CORTICOTROPIN-RELEASING FACTOR/UROTENSIN I FAMILY OF PEPTIDES.  
 CC Interpro; IPR000187; CRF.  
 DR Interpro; IPR00473; CRF; 1.  
 DR Pfam; PF00473; CRF; 1.  
 DR SMART; SM00039; CRF; 1.  
 DR PROSITE; PS00511; CRF; 1.  
 DR Hormone; Amidation.  
 KM MOD\_RES 44  
 FT MOD\_RES 44  
 SQ SEQUENCE 44 AA; 5181 MW; 67628086D21487B6 CRC64;

Query Match 15.0%; Score 38; DB 1; Length 44;  
 Best Local Similarity 25.7%; Pred. No. 6.6e+02;  
 Matches 9; Conservative 9; Mismatches 15; Indels 2; Gaps 1;  
 OY 1 PFKNKAPVSIPLDLITEAMEIKARHDEQRELE 35  
 DB 3 PSLSTVNPDLVLRQL--LLEIARRRMQSQIOANRE 35

RESULT 5  
 OSTC\_XENLA STANDARD; PRT; 49 AA.  
 AC P40147;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE OSTEOCALCIN (GAMMA-CARBOXYGLUTAMIC ACID-CONTAINING PROTEIN) (BONE GLA-PROTEIN) (BGP).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenoportinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=96139691; PubMed=8567186;  
 RA Cancelli M., Williamson M.K., Price P.A.;  
 RT "Amino-acid sequence of bone GLA protein from the African clawed toad Xenopus laevis and the fish Sparus aurata."  
 RL Int. J. Pept. Protein Res. 46:419-423(1995).  
 CC -1- FUNCTION: THIS BONE PROTEIN, CONSTITUTES 1-2% OF THE TOTAL PROTEIN OF BONE. IT BINDS STRONGLY TO APATITE AND CALCIUM.  
 CC -1- PFM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE BINDING OF CALCIUM.  
 CC -1- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN FAMILY.  
 DR Interpro; IPR002384; GLA\_bone.

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:17:42 ; Search time 13.27 Seconds  
(without alignments)  
287.018 Million cell updates/sec

Title: US-09-822-110-5

Perfect score: 253  
Sequence: 1 PIRKNAKPVSLRDLTEAM.....QRELEEEENSDDELSDHT 50

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 11081

Minimum DB seq length: 0  
Maximum DB seq length: 50

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR\_68:  
1: pirl:  
2: pirl:  
3: pirl:  
4: pirl:

'Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.'

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	19.8	49	2	A56388
2	49	19.4	41	2	T14350
3	47.5	18.8	44	2	I51889
4	47	18.6	39	2	P00011
5	46.5	18.4	50	2	F86472
6	45	17.8	44	2	C34327
7	44	17.4	38	2	I66797
8	44	17.4	43	2	S02031
9	43	17.0	48	2	S74261
10	42	16.6	45	2	A83765
11	42	16.6	46	2	T16460
12	40	15.8	40	2	A60645
13	39	15.4	44	2	I48118
14	39	15.4	46	2	A48542
15	38	15.0	25	2	S77229
16	38	15.0	36	2	JA0173
17	38	15.0	42	2	F43259
18	38	15.0	48	2	E71901
19	37.5	14.8	48	2	S61469
20	37.5	14.8	48	2	S61472
21	37	14.6	44	2	F82049
22	37	14.6	44	2	A25006
23	37	14.6	48	2	D64614
24	36.5	14.4	48	2	T07305
25	36	14.2	41	2	T48342
26	36	14.2	42	2	A39124
27	36	14.2	48	2	I46522
28	36	14.2	49	2	S08452
29	35.5	14.0	48	2	E85646

30	35.5	14.0	49	2	J01200	hypothetical 5.4k
31	35	13.8	31	2	G81558	hypothetical prote
32	35	13.8	32	2	I46523	tropomyosin T 4f - ra
33	35	13.8	34	2	S56118	myosin II heavy ch
34	35	13.8	34	2	I65746	tropomyosin - huma
35	35	13.8	42	2	T07285	hypothetical prote
36	35	13.8	45	2	T11960	ribosomal protein
37	35	13.8	48	2	S02208	osteocalcin - emu
38	35	13.8	49	2	H83820	hypothetical prote
39	35	13.8	50	2	T06541	probable NADPH-fe
40	35	13.8	50	2	S77712	mitochondrial oute
41	34	13.4	23	2	B04348	internal peptide V
42	34	13.4	23	2	A04348	internal peptide V
43	34	13.4	39	2	G82287	hypothetical prote
44	34	13.4	39	2	C83904	hypothetical prote
45	34	13.4	43	2	S41388	protein 3a - human

## ALIGNMENTS

RESULT 1  
A56388  
dopamine- and cAMP-regulated neuronal phosphoprotein DARPP-32 - rat (fragments)  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 03-Nov-1995 #sequence\_revision 03-Nov-1995 #text\_change 17-Mar-1999  
C:Accession: A56388  
R:Desdoutils, F.; Cohen, D.; Nairn, A.C.; Greengard, P.; Gilraut, J.A.  
J. Biol. Chem. 270, 8772-8778, 1995  
A:Title: Phosphorylation of DARPP-32, a dopamine- and cAMP-regulated phosphoprotein,  
A:Reference number: A56388; MUID:95236371  
A:Accession: A56388  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-49 <DE>  
A:Keywords: phosphoprotein  
F:7.35.43/Binding site: phosphate (Ser) (covalent) (by casein kinase I) #status exper

Query Match 19.8%; Score 50; DB 2; Length 49;  
Best local Similarity 37.5%; Pred. No. 1e+02; Mismatches 9; Conservative 8; Indels 0; Gaps 0;

Qy 21 EIKAKRHDEQRELEEEENSDSD 44  
DB 11 ELGYQEDDEDEDEDEDEDEED 34

RESULT 2  
T14350  
tubulin beta-3 chain - carrot (fragment)

C:Species: Daucus carota (carrot)  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Jun-2000  
C:Accession: T14350  
R:Okamura, S.; Somehara, K.; Naito, K.; Okawa, H.; Kuramori, S.; Tatsuta, M.; Minami  
submitted to the EMBL Data Library, July 1996  
A:Description: Characterization of beta-tubulin genes of carrot.

A:Reference number: Z17999  
A:Accession: T14350  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-41 <OK>  
A:Cross-references: EMBL:U64430; NID:g1553124; PTD:g1553125  
A:Experimental source: cultivar Kintoki  
C:Comment: Tubulin is a dimer of alpha and beta chains and is found in the microtubul  
changeable site on its beta chain and at a nonexchangeable site not yet identified. T  
C:Complex: heterodimer: alpha and beta chain  
C:Superfamily: tubulin  
C:Keywords: heterodimer

Query Match 19.4%; Score 49; DB 2; Length 41;  
Best local Similarity 32.4%; Pred. No. 1.1e+02;

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

C;Keywords: skeletal muscle

13 1BDITTEAMETKAKYBHDEOOREIFFFFFFFFEENSDEDEL 46

A; Authors: Hunter, J.L.; Jenkins, C.; Johnson-Hopson, S.; Kaur, G.; Kim, H.; Lin, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani

C;Keywords: skeletal muscle

13 1BDITTEAMETKAKYBHDEOOREIFFFFFFFFEENSDEDEL 46

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:05:01 ; Search time 19.84 Seconds  
(without alignments)  
907.381 Million cell updates/sec

Title: US-09-822-110-2

Perfect score: 2564  
Sequence: 1 MEGPAPAKSKLKKLSLSDSLT.....AKROPILDAMDAKKRRQNF 491

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
al number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.\*

\* Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2564	100.0	491	1	STK3_HUMAN	Q13188 homo sapien
2	1968.5	76.8	487	1	STK4_HUMAN	Q13043 homo sapien
3	817	31.9	443	1	ST24_HUMAN	O9Y660 homo sapien
4	796	31.0	426	1	ST25_HUMAN	O00506 homo sapien
5	792	30.9	426	1	ST25_MOUSE	O92241 mus musculu
6	697.5	27.2	968	1	ST10_HUMAN	O94804 homo sapien
7	693	27.0	966	1	ST10_MOUSE	O35098 mus musculu
8	634	24.7	544	1	PAK1_RAT	P33465 rattus norv
9	633	24.6	544	1	PAK1_HUMAN	O13153 homo sapien
10	630	24.6	544	1	PAK3_HUMAN	O75914 homo sapien
11	630	24.6	544	1	PAK3_MOUSE	O62829 rattus norv
12	629	24.5	544	1	PAK3_MOUSE	O61036 mus musculu
13	628.5	24.5	524	1	PAK2_RABIT	O29502 oryctolagus
14	628.5	24.5	524	1	PAK2_RAT	O64303 rattus norv
15	626.5	24.4	524	1	PAK2_HUMAN	O13177 homo sapien
16	626	24.4	545	1	PAK1_MOUSE	O88643 mus musculu
17	618.5	24.1	982	1	SIL1_CAEL	P46549 caenorhabdi
18	595.5	23.2	658	1	PAK1_SCHRO	P50527 schizosacch
19	593	23.1	1080	1	MIL5_CAEL	O23356 caenorhabdi
20	592	23.1	490	1	SPS1_YEAST	P08458 saccharomyc
21	572	22.3	971	1	CLIA4_CANAL	O14427 candida alb
22	564.5	22.0	939	1	ST20_YEAST	O03497 saccharomyc
23	554	21.6	1230	1	ST20_CANAL	O92212 candida alb
24	552	21.5	589	1	SHK2_SCHRO	O10056 schizosacch
25	552	21.5	1080	1	NKX1_YEAST	P36692 saccharomyc
26	548.5	21.4	1135	1	NKX1_DROME	P10677 drosophila
27	548.5	21.4	1501	1	NKX1_DROME	P10676 drosophila
28	533.5	20.8	547	1	SPAK_HUMAN	O94804 homo sapien
29	532.5	20.8	556	1	SPAK_MOUSE	O94804 mus musculu
30	530.5	20.7	553	1	SPAK_MOUSE	O88506 rattus norv
31	528.5	20.5	842	1	CLIA4_YEAST	P48562 saccharomyc
32	520	20.3	591	1	PAK4_HUMAN	O96013 homo sapien
33	516.5	20.1	1062	1	CC7_SCHPO	P41892 schizosacch

34	515	20.1	719	1	PAK5_HUMAN	O9P286 homo sapien
35	509.5	19.9	655	1	SKM1_YEAST	Q12469 saccharomyc
36	501	19.5	974	1	CC15_YEAST	P27636 saccharomyc
37	490	19.1	392	1	MPK1_HUMAN	O02750 homo sapien
38	490	19.1	392	1	MPK1_MOUSE	P31938 mus musculu
39	490	19.1	392	1	MPK1_RABIT	P29678 oryctolagus
40	490	19.1	392	1	MPK1_RAT	O01986 rattus norv
41	489	19.1	393	1	MPK1_CRICR	O63980 cricetus
42	486	19.0	388	1	MPK1_SERCA	O91447 serinus can
43	479	18.7	394	1	MPK1_XENLA	O05116 xenopus lae
44	478.5	18.7	401	1	MPK2_MOUSE	O63932 mus musculu
45	477	18.6	400	1	MPK2_RAT	P36506 rattus norv

## ALIGNMENTS

RESULT	ID	STK3_HUMAN	STANDARD:	PRT:	491 AA.
AC	Q13188	Q15801: Q15445;			
DT	20-AUG-2001	(Rel. 40, Created)			
DT	20-AUG-2001	(Rel. 40, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	SERINE/THREONINE PROTEIN KINASE 3 (EC 2.7.1.37) (STE20-LIKE KINASE				
DE	MST2) (MST-2) (MAMMALIAN STE20-LIKE PROTEIN KINASE 2)				
DE	(SERINE/THREONINE PROTEIN KINASE KRS-1).				
CN	STK3 OR MST2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OX	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
NC	NCBI_Taxid:9606;				
RP	SEQUENCE FROM N.A.				
RP	MDLINE=96144292; PubMed=8566796;				
RA	Creasy C.L., Chernoff J.;				
RT	"Cloning and characterization of a member of the MST subfamily of				
RT	Ste20-like kinases."				
RL	Gene 167303=306(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RP	MDLINE=96413604; PubMed=8816758;				
RA	Taylor L.K., Wang H.C., Erikson R.L.;				
RT	"Newly identified stress-responsive protein kinases, Krs-1 and Krs-				
RT	2."				
RL	Proc. Natl. Acad. Sci. U.S.A. 93:10099-10104(1996).				
RN	[3]				
RP	SEQUENCE OF 96-203 FROM N.A.				
RP	MDLINE=94100173; PubMed=8274451;				
RA	Schultz S.J., Nigg E.A.;				
RT	"Identification of 21 novel human protein kinases, including 3 members				
RT	of a family related to the cell cycle regulator NIMA of Aspergillus				
RT	nidulans."				
RL	Cell Growth Differ. 4:821-830(1993).				
CC	- FUNCTION: OXIDANT STRESS-ACTIVATED SERINE/THREONINE KINASE THAT				
CC	MAY PLAY A ROLE IN THE RESPONSE TO ENVIRONMENTAL STRESS (BY				
CC	SIMILARITY).				
CC	- CATALYTIC ACTIVITY: ATP + A PROTEIN - ADP + A PHOSPHOPROTEIN.				
CC	- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).				
CC	- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN ADULT KIDNEY,				
CC	SKLETERAL AND PLACENTA TISSUES AND AT VERY LOW LEVELS IN ADULT				
CC	HEART, LUNG AND BRAIN TISSUES.				
CC	- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.				
CC	STE20 SUBFAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				

DR EMBL: U26424; AAC50386.1;  
 DR EMBL: U60206; AAB17261.1;  
 DR EMBL: 225422; CA80909.1;  
 DR HSSP: P00518; 1PHK.  
 DR MIM: 605030;  
 DR InterPro: IPR00719; Euk\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
 DR Pfam: PF00069; kinase.1.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR SMART: SM00220; S\_TKC.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ST; FALSE\_NEG.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_DOM.1.  
 DR PROSITE: PS50011; Serine/threonine-protein kinase: ATP-binding.  
 DR Transferrase: Serine/threonine-protein kinase: ATP-binding.  
 DR TRANSFERASE: 27 278  
 FT DOMAIN 27 278  
 FT NP\_BIND 33 41  
 FT BINDING 56 56  
 FT ACT\_SITE 146 146  
 FT DOMAIN 308 314  
 FT DOMAIN 370 375  
 FT CONFLICT 96 98  
 FT CONFLICT 121 121  
 FT CONFLICT 203 203  
 FT CONFLICT 303 303  
 FT CONFLICT 332 334  
 FT SEQUENCE 491 AA; 56261 MW; 9CA3B0644F3C14A9 CRC64;

Query Match 100.0%; Score 2564; DB 1; Length 491;  
 Best Local Similarity 100.0%; Pred No. 4.5e-142;  
 Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 MEOPAPKSKLKIKLSESLTKOPEVEVDVLEKLGESYGSVFAIKHESGOVAIKOVPY 60  
 DB 1 MEOPAPKSKLKIKLSESLTKOPEVEVDVLEKLGESYGSVFAIKHESGOVAIKOVPY 60  
 DB 61 ESLOELIKRISIMOODSPYVVKYGYKNDIMVMEYCGAGSVSDIRLNKLTLE 120  
 DB 61 ESLOELIKRISIMOODSPYVVKYGYKNDIMVMEYCGAGSVSDIRLNKLTLE 120  
 QY 121 DELATILKSTLKGLEYLHFKIRHIDIKAGNILLNTEGHAKLADFCVAGOLPTMAKRW 180  
 DB 121 DELATILKSTLKGLEYLHFKIRHIDIKAGNILLNTEGHAKLADFCVAGOLPTMAKRW 180  
 QY 181 VICTPMAPEVTOELGYNCVADIMSLGITSIMAGKPPYADIHMPRAIFMIPNPPPT 240  
 DB 181 VICTPMAPEVTOELGYNCVADIMSLGITSIMAGKPPYADIHMPRAIFMIPNPPPT 240  
 QY 241 FRKPELMSDFTDFVKKLVKNPEORATATOLLOHPEIKNAKPVSLIRLITEAMEIKAK 300  
 DB 241 FRKPELMSDFTDFVKKLVKNPEORATATOLLOHPEIKNAKPVSLIRLITEAMEIKAK 300  
 QY 301 RHDEOORELEEEENSDDELDSHMVKTSGVECGTMRATSTMSGCAQTMIEHNSMTLES 360  
 DB 301 RHDEOORELEEEENSDDELDSHMVKTSGVECGTMRATSTMSGCAQTMIEHNSMTLES 360  
 QY 361 DGTWVINSDEDEEDGCTKRRATSPORPSEFMVFPKODKKNKSHENCONMHEPPPM 420  
 DB 361 DGTWVINSDEDEEDGCTKRRATSPORPSEFMVFPKODKKNKSHENCONMHEPPPM 420  
 QY 421 SKNVEPPMKNVPODGFPLKNLSLEELOMRLKALDPMMERIEELRORYAKROPILDA 480  
 DB 421 SKNVEPPMKNVPODGFPLKNLSLEELOMRLKALDPMMERIEELRORYAKROPILDA 480  
 QY 481 MDAKKRROOF 491  
 DB 481 MDAKKRROOF 491

RESULT 2  
 ID STR4\_HUMAN STANDARD;  
 AC Q13043; Q15802; Q9NT24;

PRT: 487 AA.

DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE SERINE/THREONINE PROTEIN KINASE 4 (EC 2.7.1.37) (STE20-LIKE KINASE  
 DE MST1) (MST-1) (MAMMALIAN STE20-LIKE PROTEIN KINASE 1)  
 DE (SERINE/THREONINE PROTEIN KINASE KRS-2).  
 GN SK4 OR MST1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95394929; PubMed=7665586;  
 RA Creasy C.L., Chernoff J.;  
 RT "Cloning and characterization of a human protein kinase with homology  
 RT to Ste20.";  
 RL J. Biol. Chem. 270:21695-21700(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96413604; PubMed=8816758;  
 RA Taylor L.K., Wang H.C., Erikson R.L.;  
 RT "Newly identified stress-responsive protein kinases, Krs-1 and Krs-  
 RT 2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:10099-10104(1996).  
 RN [3]  
 RP SEQUENCE OF 1-435 FROM N.A.  
 RA Latid G.;  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.  
 CC -1- ENZYME REGULATION: THE C-TERMINAL NON-CATALYTIC REGION INHIBITS  
 CC THE KINASE ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.  
 CC -1- PTM: AUTOPHOSPHORYLATED ON SERINE AND THREONINE RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC STE20 SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 76.8%; Score 1968.5; DB 1; Length 487;  
 Best Local Similarity 78.2%; Pred. No. 1.6e-107;

DR EMBL: U18297; AAA83254.1;  
 DR EMBL: U60207; AAB17262.1;  
 DR EMBL: AL109839; CAB89421.1;  
 DR HSSP: P24941; IHCL.  
 DR MIM: 604965;  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kin\_actsite.  
 DR InterPro: IPR001245; Tyr\_kin.  
 DR Pfam: PF00069; kinase.1.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR SMART: SM00220; S\_TKC.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ST; FALSE\_NEG.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_DOM.1.  
 DR PROSITE: PS50011; Serine/threonine-protein kinase: ATP-binding.  
 DR TRANSFERASE: Serine/threonine-protein kinase: ATP-binding.  
 DR TRANSFERASE: 30 281  
 FT DOMAIN 30 281  
 FT NP\_BIND 36 44  
 FT BINDING 59 59  
 FT ACT\_SITE 149 149  
 FT ACT\_SITE 373 378  
 FT DOMAIN 373 378  
 FT CONFLICT 222 222  
 FT CONFLICT 312 312  
 FT SEQUENCE 487 AA; 55630 MW; 15075B8BC5F77D5C CRC64;



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: February 15, 2002, 07:04:32 ; Search time 53.55 Seconds  
(without alignments)  
1341.171 Million cell updates/sec

Title: US-09-822-110-2  
Perfect score: 2564  
Sequence: 1 MEOPAPPSKLLKLSLSDSLT.....AKROPILDMADAKKRQNF 491

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues  
al number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_17:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_mammal:\*  
9: sp\_organella:\*  
10: sp\_phase:\*  
11: sp\_plant:\*  
12: sp\_proteus:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
15: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2462	96.0	491	11	054748
2	2442	95.2	497	11	09J110
3	2110	82.3	445	11	060877
4	1961.5	76.5	487	11	09J111
5	1555	60.6	321	11	09CW82
6	1460.5	57.0	478	5	062571
7	1411.5	51.3	497	5	09NB31
8	1315.5	42.0	461	5	061125
9	1076.5	42.0	461	5	061125
10	816.5	31.8	642	5	09VEM3
11	815.5	31.8	416	4	09BXC4
12	815.5	31.8	416	11	09J9T2
13	813	31.7	416	4	09P289
14	811.5	31.6	431	11	099KH8
15	802.5	31.3	478	5	061125
16	800	31.2	653	5	022553
17	775	30.2	1120	10	09LQ41
18	774.5	30.2	836	10	024527
19	772.5	30.1	842	10	09FNU3

20	765.5	29.9	825	10	09ARL7	09ar17 hordeum vul
21	714.5	27.9	1202	11	09W041	09w041 mus musculu
22	714.5	27.9	1233	11	054988	054988 mus musculu
23	712	27.8	690	10	09LDM6	09ldm6 arabidopsis
24	706	27.5	676	10	09ZPD6	09zpd6 brassica na
25	705.5	27.5	1615	4	09NIS8	09nis8 homo sapien
26	701.5	27.4	1218	5	09VHR6	09vhr6 drosophila
27	701.5	27.4	1231	11	055092	055092 cavia porce
28	701.5	27.4	1297	4	09UKD9	09ukd9 homo sapien
29	701.5	27.4	1305	4	09UKR3	09ukr3 homo sapien
30	701.5	27.4	1352	4	09UKR2	09ukr2 homo sapien
31	701.5	27.4	1360	4	09UKR5	09ukr5 homo sapien
32	700.5	27.3	1300	11	09JUM92	09jum92 mus musculu
33	700.5	27.3	1308	11	09JUM52	09jum52 mus musculu
34	698	27.2	1268	4	09UKR8	09ukr8 homo sapien
35	698	27.2	1276	4	09UKR1	09ukr1 homo sapien
36	698	27.2	1323	4	09UKR0	09ukr0 homo sapien
37	697.5	27.2	1331	4	09UKR4	09ukr4 homo sapien
38	697.5	27.2	1152	4	092603	092603 homo sapien
39	697.5	27.2	1204	4	000211	000211 homo sapien
40	697.5	27.2	1235	4	09H262	09h262 homo sapien
41	697	27.2	471	3	014305	014305 schizosacch
42	696	27.1	1206	11	008815	008815 rattus norv
43	694.5	27.1	1295	4	09P2R8	09p2r8 homo sapien
44	694.5	27.1	1303	4	09PIX1	09pix1 homo sapien
45	693	27.0	1165	4	095819	095819 homo sapien

## ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	491 AA.
ID	054748			
AC	054748			
DT	01-JUN-1998 (TREMBLrel. 06, Created)			
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	MSR2 KINASE.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Aurisicchio, L., DiLauro R., Zannini M.;			
RL	Submitted (Sep-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AJ001529; CAA04814.1;			
DR	HSSP; P00518; 1PHK			
DR	InterPro: IPR000719; Euk PKinase.			
DR	InterPro: IPR002290; Ser thr_kin_actsite.			
DR	Pfam; PF00069; PKinase; 1.			
DR	SMART; SM00220; S_TRC; 1.			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP_1.			
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM_1.			
KW	ATP-binding; Transferase			
SG	SEQUENCE 491 AA; 56121 MW; 90FA6B020E7FFD62 CRC64;			
Query Match	96.0%; Score 2462; DB 11; Length 491;			
Best Local Similarity	96.1%; Pred. No. 5.6e-160;			
Matches	472; Conservative 9; Mismatches 10; Indels 0; Gaps 0;			
Qy	1	MEOPAPPSKLLKLSLSDSLTROPPEVFLVLEKLGSGSVFKAIHKSGGVVAIKQVPV	60	
Db	1	MEOPAPPSKLLKLSLSDSLTROPPEVFLVLEKLGSGSVFKAIHKSGGVVAIKQVPV	60	
Qy	61	ESDLOEIKKEISIMQCCSPYVKKYGSYFKNTDLMIYMEYCGAGSVDIIRLRKLTLE	120	
Db	61	ESDVEIKKEISIMQCCSPYVKKYGSYFKNTDLMIYMEYCGAGSVDIIRLRKLTLE	120	
Qy	121	DEATILKSTLGLFYLHFMKRIRHDIKAGNILLTEGHAKLADGVAGQTLDTMAKRN	180	

```

Db 121 DEATILKSTLKGLEYLHFMKRIHRDIKAGNILLNTEGNAKLADPGVAGQLDTMAKRN 180
Qy 181 VIGTFPMMAPEVIOEIGYNCVADIMSLGITSIEMAEKGPYPADIPHMAIFMIPNPPPT 240
Db 181 VIGTFPMMAPEVIOEIGYNCVADIMSLGITSIEMAEKGPYPADIPHMAIFMIPNPPPT 240
Qy 241 FRKPELMSDDFDYKCKLVKNSPEORATATOLLOHPFKNNKPPVSLDLTAMEIKAK 300
Db 241 FRKPELMSDDFDYKCKLVKNSPEORATATOLLOHPFKNNKPPVSLDLTAMEIKAK 300
Qy 301 RHDEORELEEEENEDEDELDSHTMVKTSVSGCTMRATSTMSGAQTMIEHNSMTLES 360
Db 301 RHDEORELEEEENEDEDELDSHTMVKTSVSGCTMRATSTMSGAQTMIEHNSMTLES 360
Qy 361 DLGTWVINSDEEEEDGTMKRNATSPQVORPSFMDYFDKODFKNKNHNCNMHEPFP 420
Db 361 DLGTWVINSDEEEEDGTMKRNATSPQVORPSFMDYFDKODFKNKNHNCNMHEPFP 420
Qy 421 SKVFPDNNKVPQDGFDFLKNLSLELOMRKALDPMMERIEELRORYTAKR 480
Db 421 SKVFPDNNKVPQDGFDFLKNLSLELOMRKALDPMMERIEELRORYTAKR 480
Qy 481 MDAKKRRQNF 491
Db 481 MDAKKRRQNF 491

```

## RESULT 2

```

ID 09J110 PRELIMINARY: PRT: 497 AA.
AC 09J110;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE STE20-LIKE KINASE MST2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Lee K.-K., Ohyama T., Yajima N., Yonehara S.;
RT MST2, a physiological caspase substrate, highly sensitizes apoptosis
RT both upstream and downstream of caspase activation.
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF271361; AAF75790.1;
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR Pfam: PF000069; pkinase.1
DR PRINTS; PR00109; TYRKINASE.
DR SMART; SM00220; S_TKC.1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
KW ATP-binding; Kinase; Transferase.
SQ SEQUENCE 497 AA; 56855 MW; 9CDD365437581665 CRC64;

```

Query Match 95.2%; Score 2442; DB 11; Length 497;  
 Best Local Similarity 95.0%; Pred. No. 1.3e-158;  
 Matches 472; Conservative 8; Mismatches 11; Indels 6; Gaps 1;

```

Qy 1 MEOPPAKSKLKLSDSLTKOPEEVDVLEKLGESYGVFAIKHESQYVAIKQVPV 60
Db 1 MEOPPAKSKLKLSDSLTKOPEEVDVLEKLGESYGVFAIKHESQYVAIKQVPV 60
Qy 61 ESDLOELIKEISIMQOCDSPYVVKYGYFKNTDLMVMEYCGAGSVSDIIRLNKTLTE 120
Db 61 ESDLOELIKEISIMQOCDSPYVVKYGYFKNTDLMVMEYCGAGSVSDIIRLNKTLTE 120
Qy 121 DELATILKSTLKGLEYLHFMKRIHRDIKAGNILLNTEGNAKLADPGVAGQLDTMAKRN 180
Db 121 DELATILKSTLKGLEYLHFMKRIHRDIKAGNILLNTEGNAKLADPGVAGQLDTMAKRN 180

```

```

Db 121 DEATILKSTLKGLEYLHFMKRIHRDIKAGNILLNTEGNAKLADPGVAGQLDTMAKRN 180
Qy 181 VIGTFPMMAPEVIOEIGYNCVADIMSLGITSIEMAEKGPYPADIPHMAIFMIPNPPPT 240
Db 181 VIGTFPMMAPEVIOEIGYNCVADIMSLGITSIEMAEKGPYPADIPHMAIFMIPNPPPT 240
Qy 241 FRKPELMSDDFDYKCKLVKNSPEORATATOLLOHPFKNNKPPVSLDLTAMEIKAK 300
Db 241 FRKPELMSDDFDYKCKLVKNSPEORATATOLLOHPFKNNKPPVSLDLTAMEIKAK 300
Qy 301 RHDEORELEEEENEDEDELDSHTMVKTSVSGCTMRATSTMSGAQTMIEHNSMTLES 360
Db 301 RHDEORELEEEENEDEDELDSHTMVKTSVSGCTMRATSTMSGAQTMIEHNSMTLES 360
Qy 361 DLGTWVINSDEEEEDGTMKRNATSPQVORPSFMDYFDKODFKNKNHNCNMHEPFP 420
Db 361 DLGTWVINSDEEEEDGTMKRNATSPQVORPSFMDYFDKODFKNKNHNCNMHEPFP 420
Qy 421 SKVFPDNNKVPQDGFDFLKNLSLELOMRKALDPMMERIEELRORYTAKR 480
Db 421 SKVFPDNNKVPQDGFDFLKNLSLELOMRKALDPMMERIEELRORYTAKR 480
Qy 481 MDAKKRRQNF 491
Db 481 MDAKKRRQNF 491

```

## RESULT 3

```

ID 060877 PRELIMINARY: PRT: 445 AA.
AC 060877;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTEIN KINASE MESS1.
GN STK3 OR MESS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Han J.;
RL Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; U28726; AAA75300.1;
DR HSP; P00518; LPHK.
DR MGD; MGI:1928487; Stk3.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR Pfam: PF000069; pkinase.1.
DR SMART; SM00220; S_TKC.1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
KW ATP-binding; Transferase.
SQ SEQUENCE 445 AA; 50407 MW; 7B936B28DA616F05 CRC64;

```

Query Match 82.3%; Score 2110; DB 11; Length 445;  
 Best Local Similarity 94.9%; Pred. No. 4.9e-136;  
 Matches 409; Conservative 6; Mismatches 10; Indels 6; Gaps 1;

```

Qy 1 MEOPPAKSKLKLSDSLTKOPEEVDVLEKLGESYGVFAIKHESQYVAIKQVPV 60
Db 1 MEOPPAKSKLKLSDSLTKOPEEVDVLEKLGESYGVFAIKHESQYVAIKQVPV 60
Qy 61 ESDLOELIKEISIMQOCDSPYVVKYGYFKNTDLMVMEYCGAGSVSDIIRLNKTLTE 120
Db 61 ESDLOELIKEISIMQOCDSPYVVKYGYFKNTDLMVMEYCGAGSVSDIIRLNKTLTE 120
Qy 121 DELATILKSTLKGLEYLHFMKRIHRDIKAGNILLNTEGNAKLADPGVAGQLDTMAKRN 180
Db 121 DELATILKSTLKGLEYLHFMKRIHRDIKAGNILLNTEGNAKLADPGVAGQLDTMAKRN 180

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:03:42 ; Search time 26.93 seconds  
(without alignments)  
410.290 Million cell updates/sec

Title: US-09-822-110-2  
Perfect score: 2564  
Sequence: 1 MEQPPAPKSLKLUKLESDSLT.....AKROPILDAMDKKRRQGNF 491

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 212252 seqs, 22503292 residues

Maximum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfillsl.pep:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1960.5	76.5	487	2	US-08-712-709-8
2	1960.5	76.5	487	3	US-09-111-444-8
3	1960.5	76.5	487	4	US-09-541-228-8
4	1333	52.0	270	2	US-08-852-743-5
5	1333	52.0	270	3	US-09-185-370-5
6	817	31.9	431	2	US-09-211-930-5
7	817	31.9	431	3	US-09-340-993-5
8	817	31.9	431	4	US-09-152-406-3
9	817	31.9	431	4	US-09-468-442-5
10	815.5	31.8	416	2	US-09-211-930-3
11	815.5	31.8	416	2	US-09-211-930-11
12	815.5	31.8	416	3	US-09-340-993-3
13	815.5	31.8	416	3	US-09-340-993-11
14	815.5	31.8	416	4	US-09-468-442-3
15	815.5	31.8	416	4	US-09-468-442-11
16	793.5	30.9	426	2	US-08-852-743-2
17	793.5	30.9	426	2	US-09-211-930-4
18	793.5	30.9	426	3	US-09-340-993-4
19	793.5	30.9	426	3	US-09-185-370-2
20	793.5	30.9	426	4	US-09-152-406-4
21	793.5	30.9	426	4	US-09-468-442-4
22	739	28.8	403	2	US-08-712-709-3
23	739	28.8	403	3	US-09-111-444-3
24	739	28.8	403	4	US-09-541-228-3
25	701.5	27.4	1360	4	US-09-393-569-2
26	645	25.2	1001	4	US-09-060-410-2
27	631	24.6	465	2	US-08-114-555A-2

28	629	24.5	545	2	US-08-935-760-4	Sequence 4, App11
29	628.5	24.5	524	2	US-08-615-942A-2	Sequence 2, App11
30	628	24.5	544	3	US-08-935-760-2	Sequence 2, App11
31	627	24.5	544	3	US-08-559-397A-19	Sequence 19, App1
32	626	24.4	276	2	US-08-852-743-7	Sequence 7, App11
33	626	24.4	276	3	US-09-185-370-7	Sequence 7, App11
34	620.5	24.2	506	1	US-08-369-780-2	Sequence 2, App11
35	620.5	24.2	506	1	US-08-475-682-2	Sequence 2, App11
36	620.5	24.2	506	1	US-08-780-833-2	Sequence 2, App11
37	620.5	24.2	506	1	US-08-636-036-2	Sequence 2, App11
38	620.5	24.2	506	3	US-08-918-509-2	Sequence 2, App11
39	620.5	24.2	506	3	US-09-108-262-2	Sequence 2, App11
40	619	24.1	465	3	US-08-559-397A-2	Sequence 2, App11
41	610	23.8	268	2	US-08-852-743-3	Sequence 3, App11
42	610	23.8	268	3	US-09-185-370-3	Sequence 3, App11
43	606	23.6	544	3	US-08-559-397A-29	Sequence 29, App1
44	601	23.4	544	3	US-08-559-397A-30	Sequence 30, App1
45	598	23.3	993	4	US-09-060-410-4	Sequence 4, App11

ALIGNMENTS

RESULT 1  
US-08-712-709-8  
Sequence 8, Application US/08712709  
Patent No. 5863780  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Guegler, Karl J.  
APPLICANT: Hawkins, Phillip R.  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER-READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTA Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/712,709  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 487 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 1117791  
US-08-712-709-8  
Query Match 76.5%; Score 1960.5; DB 2: Length 487;  
Best Local Similarity 78.0%; Pred. No. 1,3e-160;  
Matches 379; Conservative 54; Mismatches 44; Indels 9; Gaps 5;  
QY 7 PKSKLKLSEDSLTKEEVEVDVLEKLGESGYSVFKAIRKESGQVVAIKQVPEVDLQ 66

```

Db 10 PRRLKRLDESLTKOPEEVDVLEKLGEGSYGVAIHKETQIYAIKQVPESDLOE 69
QY 67 IIKETSIHQCCDSPIYVYKYSYFKNTDLMVMEYCGAGSVDILRLRNKTLFDEIATI 126
Db 70 IIKETSIHQCCDSPIYVYKYSYFKNTDLMVMEYCGAGSVDILRLRNKTLFDEIATI 129
QY 127 LKSTLKGLEYLHFMKRIHRDIKAGNILLNTEGHAKLADFGVAGQLTDTMAKRNVTIGTF 186
Db 130 LQSTLKGLEYLHFMKRIHRDIKAGNILLNTEGHAKLADFGVAGQLTDTMAKRNVTIGTF 189
QY 187 WMAPEVIOEIGYNCVADWISGITSIEAEKRPYADIHPRALFEMIPTNPPFRKPEL 246
Db 190 WMAPEVIOEIGYNCVADWISGITSIEAEKRPYADIHPRALFEMIPTNPPFRKPEL 249
QY 247 WSDFTDFVKKCLVKNPQRATATQLOHPFIKNKPVSIIRDLTTEAMEIKAKRHDEQ 306
Db 250 WSDNTDFVKKCLVKNPQRATATQLOHPFIKNKPVSIIRDLTTEAMEIKAKRHDEQ 309
QY 307 REL-EEEPNSDEDELDSHTWKTSGVCGCTMRATSTSEAGQTMIEHNSMTLESIGTM 365
Db 310 REMDQDDENSEEEDMDGTWRAVAGDEMGTWRAVASTMTDANTMI EHDVT-LPSQLSTM 368
QY 366 VINSEDEEEDGTMKNRATSPVOYRSPMDYFDKODFKNKSHECNQNMHEPFPMSKNVF 425
Db 369 VINADEDEEE-GTMKRREDETQPAKPSFLEYEQ--KEKEMQINSFKSVGPGLKNS-- 422
QY 426 PDNMKVPDGDGDFLKNLSLEELQRLKALDPMMERIEIEELRQRYTAKROPILDAIAK 485
Db 423 -SDWKIPDGDGYEFLKMTVEDLQRLKALDPMMERIEIEELRQRYTAKROPILDAIAK 481
QY 486 RROQNF 491
Db 482 RROQNF 487

```

RESULT 2  
US-09-111-444-8  
Sequence 8, Application US/09111444  
Patent No. 6045792  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Guegler, Karl J.  
APPLICANT: Hawkins, Phillip R.  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/111,444  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/712,709  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0118 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-4166  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 8:  
INFORMATION CHARACTERISTICS:

```

LENGTH: 487 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1117791
US-09-111-444-8

```

Query Match 76.5%; Score 1960.5; DB 3; Length 487;  
Best Local Similarity 78.0%; Pred. No. 1,3e-160;  
Matches 379; Conservative 54; Mismatches 44; Indels 9; Gaps 5;

```

QY 7 PKSLKRLDESLTKOPEEVDVLEKLGEGSYGVAIHKETQIYAIKQVPESDLOE 66
Db 10 PRRLKRLDESLTKOPEEVDVLEKLGEGSYGVAIHKETQIYAIKQVPESDLOE 69
QY 67 IIKETSIHQCCDSPIYVYKYSYFKNTDLMVMEYCGAGSVDILRLRNKTLFDEIATI 126
Db 70 IIKETSIHQCCDSPIYVYKYSYFKNTDLMVMEYCGAGSVDILRLRNKTLFDEIATI 129
QY 127 LKSTLKGLEYLHFMKRIHRDIKAGNILLNTEGHAKLADFGVAGQLTDTMAKRNVTIGTF 186
Db 130 LQSTLKGLEYLHFMKRIHRDIKAGNILLNTEGHAKLADFGVAGQLTDTMAKRNVTIGTF 189
QY 187 WMAPEVIOEIGYNCVADWISGITSIEAEKRPYADIHPRALFEMIPTNPPFRKPEL 246
Db 190 WMAPEVIOEIGYNCVADWISGITSIEAEKRPYADIHPRALFEMIPTNPPFRKPEL 249
QY 247 WSDFTDFVKKCLVKNPQRATATQLOHPFIKNKPVSIIRDLTTEAMEIKAKRHDEQ 306
Db 250 WSDNTDFVKKCLVKNPQRATATQLOHPFIKNKPVSIIRDLTTEAMEIKAKRHDEQ 309
QY 307 REL-EEEPNSDEDELDSHTWKTSGVCGCTMRATSTSEAGQTMIEHNSMTLESIGTM 365
Db 310 REMDQDDENSEEEDMDGTWRAVAGDEMGTWRAVASTMTDANTMI EHDVT-LPSQLSTM 368
QY 366 VINSEDEEEDGTMKNRATSPVOYRSPMDYFDKODFKNKSHECNQNMHEPFPMSKNVF 425
Db 369 VINADEDEEE-GTMKRREDETQPAKPSFLEYEQ--KEKEMQINSFKSVGPGLKNS-- 422
QY 426 PDNMKVPDGDGDFLKNLSLEELQRLKALDPMMERIEIEELRQRYTAKROPILDAIAK 485
Db 423 -SDWKIPDGDGYEFLKMTVEDLQRLKALDPMMERIEIEELRQRYTAKROPILDAIAK 481
QY 486 RROQNF 491
Db 482 RROQNF 487

```

RESULT 3  
US-09-541-228-8  
Sequence 8, Application US/09541228  
Patent No. 6232077  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Guegler, Karl J.  
APPLICANT: Hawkins, Phillip R.  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASES  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:04:27 ; Search time 32.41 Seconds

(without alignments)  
37.605 Million cell updates/sec

Title: US-09-822-110-3

Perfect score: 81

Sequence: 1 EIKAKRHDEQORELEE 16

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

1 number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR:68:\*2: PIR:1:\*3: PIR:2:\*4: PIR:3:\*5: PIR:4:\*6: PIR:5:\*7: PIR:6:\*8: PIR:7:\*9: PIR:8:\*10: PIR:9:\*11: PIR:10:\*12: PIR:11:\*13: PIR:12:\*14: PIR:13:\*15: PIR:14:\*16: PIR:15:\*17: PIR:16:\*18: PIR:17:\*19: PIR:18:\*20: PIR:19:\*21: PIR:20:\*22: PIR:21:\*23: PIR:22:\*24: PIR:23:\*25: PIR:24:\*26: PIR:25:\*27: PIR:26:\*28: PIR:27:\*29: PIR:28:\*30: PIR:29:\*31: PIR:30:\*32: PIR:31:\*33: PIR:32:\*34: PIR:33:\*35: PIR:34:\*36: PIR:35:\*37: PIR:36:\*38: PIR:37:\*39: PIR:38:\*40: PIR:39:\*41: PIR:40:\*42: PIR:41:\*43: PIR:42:\*44: PIR:43:\*45: PIR:44:\*46: PIR:45:\*47: PIR:46:\*48: PIR:47:\*49: PIR:48:\*50: PIR:49:\*51: PIR:50:\*52: PIR:51:\*53: PIR:52:\*54: PIR:53:\*55: PIR:54:\*56: PIR:55:\*57: PIR:56:\*58: PIR:57:\*59: PIR:58:\*60: PIR:59:\*61: PIR:60:\*62: PIR:61:\*63: PIR:62:\*64: PIR:63:\*65: PIR:64:\*66: PIR:65:\*67: PIR:66:\*68: PIR:67:\*69: PIR:68:\*70: PIR:69:\*71: PIR:70:\*72: PIR:71:\*73: PIR:72:\*74: PIR:73:\*75: PIR:74:\*76: PIR:75:\*77: PIR:76:\*78: PIR:77:\*79: PIR:78:\*80: PIR:79:\*81: PIR:80:\*82: PIR:81:\*83: PIR:82:\*84: PIR:83:\*85: PIR:84:\*86: PIR:85:\*87: PIR:86:\*88: PIR:87:\*89: PIR:88:\*90: PIR:89:\*91: PIR:90:\*92: PIR:91:\*93: PIR:92:\*94: PIR:93:\*95: PIR:94:\*96: PIR:95:\*97: PIR:96:\*98: PIR:97:\*99: PIR:98:\*100: PIR:99:\*101: PIR:100:\*102: PIR:101:\*103: PIR:102:\*104: PIR:103:\*105: PIR:104:\*106: PIR:105:\*107: PIR:106:\*108: PIR:107:\*109: PIR:108:\*110: PIR:109:\*111: PIR:110:\*112: PIR:111:\*113: PIR:112:\*114: PIR:113:\*115: PIR:114:\*116: PIR:115:\*117: PIR:116:\*118: PIR:117:\*119: PIR:118:\*120: PIR:119:\*121: PIR:120:\*122: PIR:121:\*123: PIR:122:\*124: PIR:123:\*125: PIR:124:\*126: PIR:125:\*127: PIR:126:\*128: PIR:127:\*129: PIR:128:\*130: PIR:129:\*131: PIR:130:\*132: PIR:131:\*133: PIR:132:\*134: PIR:133:\*135: PIR:134:\*136: PIR:135:\*137: PIR:136:\*138: PIR:137:\*139: PIR:138:\*140: PIR:139:\*141: PIR:140:\*142: PIR:141:\*143: PIR:142:\*144: PIR:143:\*145: PIR:144:\*146: PIR:145:\*147: PIR:146:\*148: PIR:147:\*149: PIR:148:\*150: PIR:149:\*151: PIR:150:\*152: PIR:151:\*153: PIR:152:\*154: PIR:153:\*155: PIR:154:\*156: PIR:155:\*157: PIR:156:\*158: PIR:157:\*159: PIR:158:\*160: PIR:159:\*161: PIR:160:\*162: PIR:161:\*163: PIR:162:\*164: PIR:163:\*165: PIR:164:\*166: PIR:165:\*167: PIR:166:\*168: PIR:167:\*169: PIR:168:\*170: PIR:169:\*171: PIR:170:\*172: PIR:171:\*173: PIR:172:\*174: PIR:173:\*175: PIR:174:\*176: PIR:175:\*177: PIR:176:\*178: PIR:177:\*179: PIR:178:\*180: PIR:179:\*181: PIR:180:\*182: PIR:181:\*183: PIR:182:\*184: PIR:183:\*185: PIR:184:\*186: PIR:185:\*187: PIR:186:\*188: PIR:187:\*189: PIR:188:\*190: PIR:189:\*191: PIR:190:\*192: PIR:191:\*193: PIR:192:\*194: PIR:193:\*195: PIR:194:\*196: PIR:195:\*197: PIR:196:\*198: PIR:197:\*199: PIR:198:\*200: PIR:199:\*201: PIR:200:\*202: PIR:201:\*203: PIR:202:\*204: PIR:203:\*205: PIR:204:\*206: PIR:205:\*207: PIR:206:\*208: PIR:207:\*209: PIR:208:\*210: PIR:209:\*211: PIR:210:\*212: PIR:211:\*213: PIR:212:\*214: PIR:213:\*215: PIR:214:\*216: PIR:215:\*217: PIR:216:\*218: PIR:217:\*219: PIR:218:\*220: PIR:219:\*221: PIR:220:\*222: PIR:221:\*223: PIR:222:\*224: PIR:223:\*225: PIR:224:\*226: PIR:225:\*227: PIR:226:\*228: PIR:227:\*229: PIR:228:\*230: PIR:229:\*231: PIR:230:\*232: PIR:231:\*233: PIR:232:\*234: PIR:233:\*235: PIR:234:\*236: PIR:235:\*237: PIR:236:\*238: PIR:237:\*239: PIR:238:\*240: PIR:239:\*241: PIR:240:\*242: PIR:241:\*243: PIR:242:\*244: PIR:243:\*245: PIR:244:\*246: PIR:245:\*247: PIR:246:\*248: PIR:247:\*249: PIR:248:\*250: PIR:249:\*251: PIR:250:\*252: PIR:251:\*253: PIR:252:\*254: PIR:253:\*255: PIR:254:\*256: PIR:255:\*257: PIR:256:\*258: PIR:257:\*259: PIR:258:\*260: PIR:259:\*261: PIR:260:\*262: PIR:261:\*263: PIR:262:\*264: PIR:263:\*265: PIR:264:\*266: PIR:265:\*267: PIR:266:\*268: PIR:267:\*269: PIR:268:\*270: PIR:269:\*271: PIR:270:\*272: PIR:271:\*273: PIR:272:\*274: PIR:273:\*275: PIR:274:\*276: PIR:275:\*277: PIR:276:\*278: PIR:277:\*279: PIR:278:\*280: PIR:279:\*281: PIR:280:\*282: PIR:281:\*283: PIR:282:\*284: PIR:283:\*285: PIR:284:\*286: PIR:285:\*287: PIR:286:\*288: PIR:287:\*289: PIR:288:\*290: PIR:289:\*291: PIR:290:\*292: PIR:291:\*293: PIR:292:\*294: PIR:293:\*295: PIR:294:\*296: PIR:295:\*297: PIR:296:\*298: PIR:297:\*299: PIR:298:\*300: PIR:299:\*301: PIR:300:\*302: PIR:301:\*303: PIR:302:\*304: PIR:303:\*305: PIR:304:\*306: PIR:305:\*307: PIR:306:\*308: PIR:307:\*309: PIR:308:\*310: PIR:309:\*311: PIR:310:\*312: PIR:311:\*313: PIR:312:\*314: PIR:313:\*315: PIR:314:\*316: PIR:315:\*317: PIR:316:\*318: PIR:317:\*319: PIR:318:\*320: PIR:319:\*321: PIR:320:\*322: PIR:321:\*323: PIR:322:\*324: PIR:323:\*325: PIR:324:\*326: PIR:325:\*327: PIR:326:\*328: PIR:327:\*329: PIR:328:\*330: PIR:329:\*331: PIR:330:\*332: PIR:331:\*333: PIR:332:\*334: PIR:333:\*335: PIR:334:\*336: PIR:335:\*337: PIR:336:\*338: PIR:337:\*339: PIR:338:\*340: PIR:339:\*341: PIR:340:\*342: PIR:341:\*343: PIR:342:\*344: PIR:343:\*345: PIR:344:\*346: PIR:345:\*347: PIR:346:\*348: PIR:347:\*349: PIR:348:\*350: PIR:349:\*351: PIR:350:\*352: PIR:351:\*353: PIR:352:\*354: PIR:353:\*355: PIR:354:\*356: PIR:355:\*357: PIR:356:\*358: PIR:357:\*359: PIR:358:\*360: PIR:359:\*361: PIR:360:\*362: PIR:361:\*363: PIR:362:\*364: PIR:363:\*365: PIR:364:\*366: PIR:365:\*367: PIR:366:\*368: PIR:367:\*369: PIR:368:\*370: PIR:369:\*371: PIR:370:\*372: PIR:371:\*373: PIR:372:\*374: PIR:373:\*375: PIR:374:\*376: PIR:375:\*377: PIR:376:\*378: PIR:377:\*379: PIR:378:\*380: PIR:379:\*381: PIR:380:\*382: PIR:381:\*383: PIR:382:\*384: PIR:383:\*385: PIR:384:\*386: PIR:385:\*387: PIR:386:\*388: PIR:387:\*389: PIR:388:\*390: PIR:389:\*391: PIR:390:\*392: PIR:391:\*393: PIR:392:\*394: PIR:393:\*395: PIR:394:\*396: PIR:395:\*397: PIR:396:\*398: PIR:397:\*399: PIR:398:\*400: PIR:399:\*401: PIR:400:\*402: PIR:401:\*403: PIR:402:\*404: PIR:403:\*405: PIR:404:\*406: PIR:405:\*407: PIR:406:\*408: PIR:407:\*409: PIR:408:\*410: PIR:409:\*411: PIR:410:\*412: PIR:411:\*413: PIR:412:\*414: PIR:413:\*415: PIR:414:\*416: PIR:415:\*417: PIR:416:\*418: PIR:417:\*419: PIR:418:\*420: PIR:419:\*421: PIR:420:\*422: PIR:421:\*423: PIR:422:\*424: PIR:423:\*425: PIR:424:\*426: PIR:425:\*427: PIR:426:\*428: PIR:427:\*429: PIR:428:\*430: PIR:429:\*431: PIR:430:\*432: PIR:431:\*433: PIR:432:\*434: PIR:433:\*435: PIR:434:\*436: PIR:435:\*437: PIR:436:\*438: PIR:437:\*439: PIR:438:\*440: PIR:439:\*441: PIR:440:\*442: PIR:441:\*443: PIR:442:\*444: PIR:443:\*445: PIR:444:\*446: PIR:445:\*447: PIR:446:\*448: PIR:447:\*449: PIR:448:\*450: PIR:449:\*451: PIR:450:\*452: PIR:451:\*453: PIR:452:\*454: PIR:453:\*455: PIR:454:\*456: PIR:455:\*457: PIR:456:\*458: PIR:457:\*459: PIR:458:\*460: PIR:459:\*461: PIR:460:\*462: PIR:461:\*463: PIR:462:\*464: PIR:463:\*465: PIR:464:\*466: PIR:465:\*467: PIR:466:\*468: PIR:467:\*469: PIR:468:\*470: PIR:469:\*471: PIR:470:\*472: PIR:471:\*473: PIR:472:\*474: PIR:473:\*475: PIR:474:\*476: PIR:475:\*477: PIR:476:\*478: PIR:477:\*479: PIR:478:\*480: PIR:479:\*481: PIR:480:\*482: PIR:481:\*483: PIR:482:\*484: PIR:483:\*485: PIR:484:\*486: PIR:485:\*487: PIR:486:\*488: PIR:487:\*489: PIR:488:\*490: PIR:489:\*491: PIR:490:\*492: PIR:491:\*493: PIR:492:\*494: PIR:493:\*495: PIR:494:\*496: PIR:495:\*497: PIR:496:\*498: PIR:497:\*499: PIR:498:\*500: PIR:499:\*501: PIR:500:\*502: PIR:501:\*503: PIR:502:\*504: PIR:503:\*505: PIR:504:\*506: PIR:505:\*507: PIR:506:\*508: PIR:507:\*509: PIR:508:\*510: PIR:509:\*511: PIR:510:\*512: PIR:511:\*513: PIR:512:\*514: PIR:513:\*515: PIR:514:\*516: PIR:515:\*517: PIR:516:\*518: PIR:517:\*519: PIR:518:\*520: PIR:519:\*521: PIR:520:\*522: PIR:521:\*523: PIR:522:\*524: PIR:523:\*525: PIR:524:\*526: PIR:525:\*527: PIR:526:\*528: PIR:527:\*529: PIR:528:\*530: PIR:529:\*531: PIR:530:\*532: PIR:531:\*533: PIR:532:\*534: PIR:533:\*535: PIR:534:\*536: PIR:535:\*537: PIR:536:\*538: PIR:537:\*539: PIR:538:\*540: PIR:539:\*541: PIR:540:\*542: PIR:541:\*543: PIR:542:\*544: PIR:543:\*545: PIR:544:\*546: PIR:545:\*547: PIR:546:\*548: PIR:547:\*549: PIR:548:\*550: PIR:549:\*551: PIR:550:\*552: PIR:551:\*553: PIR:552:\*554: PIR:553:\*555: PIR:554:\*556: PIR:555:\*557: PIR:556:\*558: PIR:557:\*559: PIR:558:\*560: PIR:559:\*561: PIR:560:\*562: PIR:561:\*563: PIR:562:\*564: PIR:563:\*565: PIR:564:\*566: PIR:565:\*567: PIR:566:\*568: PIR:567:\*569: PIR:568:\*570: PIR:569:\*571: PIR:570:\*572: PIR:571:\*573: PIR:572:\*574: PIR:573:\*575: PIR:574:\*576: PIR:575:\*577: PIR:576:\*578: PIR:577:\*579: PIR:578:\*580: PIR:579:\*581: PIR:580:\*582: PIR:581:\*583: PIR:582:\*584: PIR:583:\*585: PIR:584:\*586: PIR:585:\*587: PIR:586:\*588: PIR:587:\*589: PIR:588:\*590: PIR:589:\*591: PIR:590:\*592: PIR:591:\*593: PIR:592:\*594: PIR:593:\*595: PIR:594:\*596: PIR:595:\*597: PIR:596:\*598: PIR:597:\*599: PIR:598:\*600: PIR:599:\*601: PIR:600:\*602: PIR:601:\*603: PIR:602:\*604: PIR:603:\*605: PIR:604:\*606: PIR:605:\*607: PIR:606:\*608: PIR:607:\*609: PIR:608:\*610: PIR:609:\*611: PIR:610:\*612: PIR:611:\*613: PIR:612:\*614: PIR:613:\*615: PIR:614:\*616: PIR:615:\*617: PIR:616:\*618: PIR:617:\*619: PIR:618:\*620: PIR:619:\*621: PIR:620:\*622: PIR:621:\*623: PIR:622:\*624: PIR:623:\*625: PIR:624:\*626: PIR:625:\*627: PIR:626:\*628: PIR:627:\*629: PIR:628:\*630: PIR:629:\*631: PIR:630:\*632: PIR:631:\*633: PIR:632:\*634: PIR:633:\*635: PIR:634:\*636: PIR:635:\*637: PIR:636:\*638: PIR:637:\*639: PIR:638:\*640: PIR:639:\*641: PIR:640:\*642: PIR:641:\*643: PIR:642:\*644: PIR:643:\*645: PIR:644:\*646: PIR:645:\*647: PIR:646:\*648: PIR:647:\*649: PIR:648:\*650: PIR:649:\*651: PIR:650:\*652: PIR:651:\*653: PIR:652:\*654: PIR:653:\*655: PIR:654:\*656: PIR:655:\*657: PIR:656:\*658: PIR:657:\*659: PIR:658:\*660: PIR:659:\*661: PIR:660:\*662: PIR:661:\*663: PIR:662:\*664: PIR:663:\*665: PIR:664:\*666: PIR:665:\*667: PIR:666:\*668: PIR:667:\*669: PIR:668:\*670: PIR:669:\*671: PIR:670:\*672: PIR:671:\*673: PIR:672:\*674: PIR:673:\*675: PIR:674:\*676: PIR:675:\*677: PIR:676:\*678: PIR:677:\*679: PIR:678:\*680: PIR:679:\*681: PIR:680:\*682: PIR:681:\*683: PIR:682:\*684: PIR:683:\*685: PIR:684:\*686: PIR:685:\*687: PIR:686:\*688: PIR:687:\*689: PIR:688:\*690: PIR:689:\*691: PIR:690:\*692: PIR:691:\*693: PIR:692:\*694: PIR:693:\*695: PIR:694:\*696: PIR:695:\*697: PIR:696:\*698: PIR:697:\*699: PIR:698:\*700: PIR:699:\*701: PIR:700:\*702: PIR:701:\*703: PIR:702:\*704: PIR:703:\*705: PIR:704:\*706: PIR:705:\*707: PIR:706:\*708: PIR:707:\*709: PIR:708:\*710: PIR:709:\*711: PIR:710:\*712: PIR:711:\*713: PIR:712:\*714: PIR:713:\*715: PIR:714:\*716: PIR:715:\*717: PIR:716:\*718: PIR:717:\*719: PIR:718:\*720: PIR:719:\*721: PIR:720:\*722: PIR:721:\*723: PIR:722:\*724: PIR:723:\*725: PIR:724:\*726: PIR:725:\*727: PIR:726:\*728: PIR:727:\*729: PIR:728:\*730: PIR:729:\*731: PIR:730:\*732: PIR:731:\*733: PIR:732:\*734: PIR:733:\*735: PIR:734:\*736: PIR:735:\*737: PIR:736:\*738: PIR:737:\*739: PIR:738:\*740: PIR:739:\*741: PIR:740:\*742: PIR:741:\*743: PIR:742:\*744: PIR:743:\*745: PIR:744:\*746: PIR:745:\*747: PIR:746:\*748: PIR:747:\*749: PIR:748:\*750: PIR:749:\*751: PIR:750:\*752: PIR:751:\*753: PIR:752:\*754: PIR:753:\*755: PIR:754:\*756: PIR:755:\*757: PIR:756:\*758: PIR:757:\*759: PIR:758:\*760: PIR:759:\*761: PIR:760:\*762: PIR:761:\*763: PIR:762:\*764: PIR:763:\*765: PIR:764:\*766: PIR:765:\*767: PIR:766:\*768: PIR:767:\*769: PIR:768:\*770: PIR:769:\*771: PIR:770:\*772: PIR:771:\*773: PIR:772:\*774: PIR:773:\*775: PIR:774:\*776: PIR:775:\*777: PIR:776:\*778: PIR:777:\*779: PIR:778:\*780: PIR:779:\*781: PIR:780:\*782: PIR:781:\*783: PIR:782:\*784: PIR:783:\*785: PIR:784:\*786: PIR:785:\*787: PIR:786:\*788: PIR:787:\*789: PIR:788:\*790: PIR:789:\*791: PIR:790:\*792: PIR:791:\*793: PIR:792:\*794: PIR:793:\*795: PIR:794:\*796: PIR:795:\*797: PIR:796:\*798: PIR:797:\*799: PIR:798:\*800: PIR:799:\*801: PIR:800:\*802: PIR:801:\*803: PIR:802:\*804: PIR:803:\*805: PIR:804:\*806: PIR:805:\*807: PIR:806:\*808: PIR:807:\*809: PIR:808:\*810: PIR:809:\*811: PIR:810:\*812: PIR:811:\*813: PIR:812:\*814: PIR:813:\*815: PIR:814:\*816: PIR:815:\*817: PIR:816:\*818: PIR:817:\*819: PIR:818:\*820: PIR:819:\*821: PIR:820:\*822: PIR:821:\*823: PIR:822:\*824: PIR:823:\*825: PIR:824:\*826: PIR:825:\*827: PIR:826:\*828: PIR:827:\*829: PIR:828:\*830: PIR:829:\*831: PIR:830:\*832: PIR:831:\*833: PIR:832:\*834: PIR:833:\*835: PIR:834:\*836: PIR:8

Query Match	55.68;	Score 45;	DB 2;	Length 1427;
Best Local Similarity	58.38;	Pred. No. 83;		
Matches 7: Conservative	5;	Mismatches	0;	Indels 0;
				Gaps 0;

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2002, 07:03:42 ; Search time 54.15 Seconds

(without alignments)  
671.653 Million cell updates/sec

Title: US-09-822-110-2

Perfect score: 2564  
Sequence: 1 MEOPAPKSKLKLSLSESLT.....AKROPILDMADAKKRRQNF 491

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

al number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: A\_Geneseq\_1101.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1957.5	76.3	487	AAV21674	Human Ste20 homolo
2	817	31.9	431	AAV55950	Human MEK3 protein
3	817	31.9	431	AAV04473	Human STE20-like s
4	817	31.9	431	AAV82276	Human STE20-like k
5	817	31.9	431	AAV68771	Amino acid sequenc
6	817	31.9	431	AAV87069	Human polypeptide
7	815.5	31.8	416	AAV55927	Human SIK2 protei
8	815.5	31.8	416	AAV22648	A human signal tra
9	815.5	31.8	416	AAV22651	A murine signal tr
10	815.5	31.8	416	AAV21673	Murine 5e.new poly
11	815.5	31.8	416	AAV82274	Human Ste20-like s

12	815.5	31.8	416	AAV82277	Mouse protein sequ
13	815.5	31.8	416	AAV40348	Human polypeptide
14	815.5	31.8	416	AAV71958	Human TGF-beta rec
15	815.5	31.8	416	AAV66607	Human h2252 protei
16	814.5	31.8	413	AAV21672	Murine C12.2bs pol
17	800	31.2	443	AAV55951	Nematode STE20-rel
18	796	31.0	426	AAV55949	Human STE20 protei
19	793.5	30.9	426	AAV82275	Human MAPK-pathway
20	789	30.8	426	AAV21603	Human protein kin
21	739	28.8	403	AAV54024	Human protein kin
22	714.5	27.9	1202	AAV808521	A murine Ste20-rel
23	705.5	27.5	1615	AAV03536	Human protein kin
24	701.5	27.4	1298	AAV88222	Amino acid sequenc
25	701.5	27.4	1306	AAV88218	Amino acid sequenc
26	701.5	27.4	1353	AAV88219	Amino acid sequenc
27	701.5	27.4	1360	AAV85263	Human protein kin
28	701.5	27.4	1360	AAV80059	Large NIK-related
29	698	27.2	1269	AAV88223	Amino acid sequenc
30	698	27.2	1277	AAV88220	Amino acid sequenc
31	698	27.2	1324	AAV88221	Amino acid sequenc
32	698	27.2	1332	AAV88217	Amino acid sequenc
33	697.5	27.2	912	AAV55939	Human GSK2 protein
34	697.5	27.2	968	AAV55966	Full length human
35	693	27.0	911	AAV55957	Mouse STE20-relat
36	693	27.0	1212	AAV5368	Human kinase (PKIN
37	693	27.0	1239	AAV55931	Human ZC1 protein.
38	692	27.0	1165	AAV43016	Human ORF2780
39	691	27.0	1135	AAV68784	Amino acid sequenc
40	678.5	26.5	894	AAV55935	Human KHS2 protein
41	666	26.0	842	AAV55956	Human STE20-relat
42	664.5	25.9	1233	AAV55954	Mouse STE20-relat
43	660	25.7	1106	AAV03552	Human protein kin
44	649	25.3	1297	AAV55932	Human ZC2 protein.
45	645	25.2	1001	AAV49896	Rat TA01 kinase.

## ALIGNMENTS

RESULT 1	AAV21674	AAV21674 standard; Protein: 487 AA.
XX	AAV21674	
AC	AAV21674	
XX		
DT	18-AUG-1999	(first entry)
XX		
DE	Human Ste20 homologue polypeptide S201.	
XX		
KW	Sterile 20; Ste20; serine/threonine kinase; C12.2bs; 5e.new; murine;	
KW	erythroleukemia; mitogenic signaling; cell differentiation; metastasis;	
KW	signaling pathway; cancer; ischemic stroke; heart disease; inflammation.	
XX		
OS	Homo sapiens.	
XX		
PN	MO9929857-A1.	
XX		
PD	17-JUN-1999	
XX		
PF	09-DEC-1998;	98WO-US26116.
XX		
PR	09-DEC-1997;	97US-0069078.
XX		
PA	(CHIL-) CHILDRENS MEDICAL CENT.	
XX		
PI	Agarwal S, Best J, Vail B, Zon Li	
XX		
DR	WPI; 1999-385605/32.	
XX		
PT	Murine homologues of the Sterile 20 family of serine/threonine	
XX	kinases, useful in elucidation of mitogenic signaling pathways	
PS	Disclosure; Fig 4; 37pp; English.	

XX The invention provides two members of the Sterile 20 (Ste20) family of  
 CC serine/threonine kinases (C12.2bs and 5e.new). Host cells transfected  
 CC with vectors comprising the nucleic acids encoding the polypeptides are  
 CC used for their recombinant expression. The Ste20 kinases, isolated from a  
 CC murine erythroleukemia cDNA library, are regulatory molecules involved in  
 CC mitogenic signaling as well as other cellular phenomena such as  
 CC morphology and motility. These activities are important factors in  
 CC development, cell differentiation, cancer and metastases. Manipulation of  
 CC the C12.2bs and 5e.new proteins or polynucleotides are useful for  
 CC manipulation of the signaling pathways involved and will allow  
 CC development of reagents to modulate these signaling pathways. The kinases  
 CC are also important for ischemic stroke, heart disease and inflammation.  
 CC The antibodies can be used to inhibit the activity of the kinase,  
 CC especially in vitro and in cell extracts, as well as identify the kinases  
 CC in immunosays. The present sequence represents a 5201 polypeptide, a  
 CC human homologue of Ste20.

XX Sequence 487 AA:

Query Match 76.3%; Score 1957.5; DB 20; Length 487;  
 Best Local Similarity 77.2%; Pred. No. 2.2e-161; Indels 19; Gaps 5;  
 Matches 379; Conservative 54; Mismatches 39;

QY 7 PKSKIKKISDSLTQPEEVEFDVLEKLGSGYSVFKAIHKGSGOVVAIKQVSVESDIOE 66  
 Db 10 pqrqkkkldedstlkqpeevfaylekligsgysvfkahkhetgvtalkqyvsdldqe 69  
 QY 67 IIRKISIMQCDSPYVYKYSYFKNTDVIWYEGCAGSVSPITRLNKTLEDEIARI 126  
 Db 70 IIRKISIMQCDSPYVYKYSYFKNTDVIWYEGCAGSVSPITRLNKTLEDEIARI 129  
 QY 127 LKSTLKGLLEYLHFMKRIHRIKAGNILLNTEGNAKLADFGVAGQLTDMAKRNTVGTPT 186  
 Db 130 Iqstklqleylhmrglhrdqkgnllntegnakladfgvaqqltdmakrntvgtptf 189  
 QY 187 WMAEVIQIEIGYNCVADISLIGTISIMAGSKRPYADIIHMRALFEMIPPTFFKRPPL 246  
 Db 190 WMAEVIQIEIGYNCVADISLIGTISIMAGSKRPYADIIHMRALFEMIPPTFFKRPPL 249  
 QY 247 WSDDTDVKKKCLVKNPQFQATATQLOHPFIKNAKPVSLRDLITAMEIKAKRHDEQ 306  
 Db 250 WSDTDMVTVKGLVKSPPQRTATQLOHPFIKNAKPVSLRDLITAMEIKAKRHDEQ 309  
 QY 307 REL-EEBENSDEDELDSHTWKTSGEGCTMRATSTMSBQAQTMIEHNSMLESDIGTM 365  
 Db 310 revdgddeenseedemdsqtmvrtagdemgtvvaastmsgganumlehgdt-ipsqqlgm 368  
 QY 366 VINSDEDEFEQDGTMRNTPQVOPSPMDYFDKQDFENKSHENCONMHEPFPMSKNVF 425  
 Db 369 Vintedeeeee.gtmkrirdeumpraksfleyfeqekenglns-----fgknvs 416  
 QY 426 -----PDNKKVPOQDGFDFLKNLSLEELQMRALKALDPMERIELEIKORTAKRQPLIDA 480  
 Db 417 gsklssstkwkkipqgdyelekwtvedlqkrlaldpmemqemeeirgkystrqplida 476  
 QY 481 MDAKRRROONF 491  
 Db 477 Ieakkrtrgnf 487

RESULT 2  
 AAY55950 standard; protein; 431 AA.  
 ID AAY55950:  
 XX AAY55950:  
 AC AAY55950:  
 XX 18-FEB-2000 (first entry)  
 DT 18-FEB-2000 (first entry)  
 XX Human MST3 protein.  
 DE Human MST3 protein.  
 XX Antihumantic; antitubercitic; antinflammatory; antiallergic; osteopathic;  
 KW

KW antiporiatic; antitubercitic; antistimatic; immunosuppressive;  
 KW neuroprotective; cardiatic; cerebroprotective; cytosstatic; antidiabetic;  
 KW vulnery; STE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;  
 KW ZC1, ZC2, ZC3, ZC4, KHS2, SU01, SU03, GSK2, PAK4, PAK5; antagonist;  
 KW antibody; gene therapy; rheumatoid arthritis; atherosclerosis; asthma;  
 KW inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;  
 KW rhinitis; autoimmunity; organ transplantation; multiple sclerosis;  
 KW myocardial-infarction; cardiovascular disease; stroke; renal failure;  
 KW oxidative stress-related neurodegenerative disorder; Parkinson's disease;  
 KW amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;  
 KW ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;  
 KW mesangial disorder; growth regulation; wound healing; T cell activation;  
 KW immunosuppressant.

OS Homo sapiens.  
 PN WO9953036-A2.  
 PD 21-OCT-1999.  
 PF 13-APR-1999; 99WO-US08150.  
 PR 14-APR-1998; 98US-0081784.  
 PA (SUCE-) SUGEN INC.  
 PI Plowman G, Martinez R, Whyte D;  
 XX WPI; 1999-611301/52.  
 XX Novel kinase-related polypeptides used for the diagnosis and treatment  
 XX of kinase-related diseases and disorders  
 XX Disclosure; Page 330-332; 387pp; English.

XX This sequence represents a novel STE20-related protein kinase. The  
 CC invention relates to nucleic acid molecule encoding a kinase polypeptide  
 CC selected from STLK2, STLK3, STLK4, STLK5, STLK6, STLK7, ZC1, ZC2, ZC3,  
 CC ZC4, KHS2, SU01, SU03, GSK2, PAK4 and PAK5. The proteins are used to  
 CC identify agonists and antagonists, and to raise antibodies. The  
 CC polynucleotides are useful in gene therapy protocols. The polynucleotides,  
 CC polypeptides, antibodies, antagonists and agonists may be used to treat  
 CC diseases such as immune-related disorders and diseases (e.g., rheumatoid  
 CC arthritis, atherosclerosis, chronic inflammatory bowel disease (e.g.,  
 CC Crohn's disease), asthma, osteoarthritis, psoriasis, atherosclerosis,  
 CC rhinitis, autoimmunity, and organ transplantation, chronic inflammatory  
 CC pelvic disease, multiple sclerosis, organ transplantation, myocardial  
 CC infarction, cardiovascular disease, stroke, renal failure, oxidative  
 CC stress-related neurodegenerative disorders (e.g., amyotrophic lateral  
 CC sclerosis, Parkinson's disease and Leigh syndrome), cancer,  
 CC cardiomyopathies, ischemic disorders, inflammatory disorders, diabetes  
 CC mellitus, fibrotic and mesangial disorders. The proteins may also be  
 CC useful for cell growth regulation (e.g., in wound healing), T cell  
 CC activation, mitosis control, and as immunosuppressants.

XX Sequence 431 AA:

Query Match 31.9%; Score 817; DB 20; Length 431;  
 Best Local Similarity 39.3%; Pred. No. 1.6e-62;  
 Matches 188; Conservative 73; Mismatches 123; Indels 94; Gaps 13;

QY 18 SLTRQPEEVEFDVLEKLGSGYSVFKAIHKGSGOVVAIKQVSVESDIOEIIKISIM 74  
 Db 15 nlkadepeelftklekigksgfgevfglnrqlkvaaalkldaeadeedidqgetvl 74  
 QY 75 QOCDSPYVYKYSYFKNTDVIWYEGCAGSVSPITRLNKTLEDEIATITKSTLGL 134  
 Db 75 sqcdspvrvtkyysylkdklwlimeylgggsaldl--leppldetqatallreilgyl 132  
 QY 135 EYLHFMKRIHRIKAGNILLNTEGNAKLADFGVAGQLTDMAKRNTVGTPTFMMAPVIO 194  
 Db 133 dylhsekklhrdikaanvlllsehgvekladfgvaqqltdtqlkrntvgtptfmmapevkl 192



[illegible]

30	556	21.9	622	2	T15467	hypothetical protein
31	554	21.6	1230	2	T18256	probable serine/thr
32	554	21.6	1230	2	T18259	serine/threonine p
33	552	21.5	589	2	T38086	serine/threonine-p
34	552	21.5	1080	2	S48944	hypothetical prote
35	549.5	21.4	1135	1	A29813	132k nlnac protein
36	549.5	21.4	1501	1	B29813	174k nlnac protein
37	547	21.3	353	2	T01479	hypothetical prote
38	540	21.1	561	2	T51417	protein kinase-1lk
39	538	21.0	693	2	B85112	hypothetical prote
40	531	20.7	108	2	T138212	protein-serine/thr
41	526.5	20.5	842	2	S60402	protein kinase CLA
42	520	20.3	607	2	T01904	hypothetical prote
43	518.5	20.2	1014	2	T31109	myosin III - Atlan
44	516.5	20.1	1062	2	S46367	protein kinase CDO
45	509.5	19.9	655	2	S51884	probable protein k

## ALIGNMENTS

RESULT 1  
T34356  
hypothetical protein T19A5.2 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34356  
R:Bradshaw, H.  
submitted to the EMBL Data Library, March 1996  
A:Description: The sequence of *C. elegans* cosmid T19A5.  
A:Reference number: Z21512  
A:Accession: T34356  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-653 <BBA>  
A:CROSS-references: EMBL:U053153; PIDB:MAC69038.1; GSPDB:GN00023; CESP:T19A5.2  
A:Experimental source: strain Bristol N2; clone T19A5  
C:Genetics:  
A:Gene: CESP:T19A5.2  
A:Map position: 5  
A:Introns: 26/3; 101/3; 157/1; 209/3; 272/3; 305/2; 373/3; 635/3

query match	31.28;	Score 800;	DB 2;	Length 653;
-------------	--------	------------	-------	-------------

Best Local Similarity 41.38; Pred. No. 3.1e-24;  
Matches 163; Conservative 68; Mismatches 102; Indels 62; Gaps 6

[illegible]

Db 342 -----DDGGTVRQRTDPRRAQ 357

# RESULT 2

Ste20-like protein kinase - human

C:Species: Homo sapiens (man)

C:Date: 29-Jan-1998 #sequence\_revision 06-Feb-1998 #text\_change 21-Jul-2000

C:Accession: S71886; G01022

R:PubMed, C.M.; Bonventre, J.V.; Molnar, A.; Kyriakis, J.; Force, T.

EMBO J. 15, 4537-4546, 1996

A:Title: Activation of a human Ste20-like kinase by oxidant stress defines a novel stress

A:Reference number: S71886; M01D:97042345

A:Accession: S71886

A:Molecule type: mRNA

A:Residues: 1-426 <POM>

A:Cross-references: EMBL:X99325; NID:q1430821; PIDN:CAA67700.1; PID:q1430822

A:Experimental source: cell type B cell

A:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

A:Keywords: ATP; autophosphorylation; phosphotransferase; protein kinase; stress-induced

F18-270/domain: protein kinase homology <kin>

Query Match 30.9%; Score 793.5; DB 2; Length 426;

Best Local Similarity 52.2%; Pred. No. 3.8e-24;

Matches 157; Conservative 53; Mismatches 78; Indels 13; Gaps 5;

QY 23 PEEFVDLEKIGEGSYGSVERKAHKEGGOVAIKQV--ESDLEIKETISIMQCCDS 79

Db 16 PEEFTKIDRIKSGSFEYVKGIDNHKEVVAIKIIDELEADEIDIOETLVLSQCCS 75

QY 80 PYVVKYIGSYRKNDLIMVWEYCGAGSVSDILRNKLTIEDIATILKSTLGLLEYLHF 139

Db 76 PYIRTEFSSYKSKYKIMTYELGGSSALD--LKPGLEEVYIATILRELKGLDYLSH 133

QY 140 MKRIHDKAGNINILNTEGHAKLADFGVAGQLDFTMAKRNVTIGTPPMAPVYIOELGYN 199

Db 134 ERKIHDKANVILNTEGHAKLADFGVAGQLDFTMAKRNVTIGTPPMAPVYIOELGYN 193

QY 200 CVADISLIGITISIMAEKRPYADIHPRATFMTPTNPPTFRPELMSDFTDVEYKCL 259

Db 194 FKADISLIGITISIMAEKRPYADIHPRATFMTPTNPPTFRPELMSDFTDVEYKCL 251

QY 260 VKNPEGRATQLOHPFI-KNAKPVSLRLDITAMEITAKRDEOORELEEEENSDE 318

Db 252 NKDPRRPFAKELKHKFTYTKTSFLTELIDRYKMKSEGHGE-----ESSSESDI 306

QY 319 D 319

Db 307 D 307

# RESULT 3

probable serine/threonine kinase F23010.20 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: B96716

R:PubMed, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Liu, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzall,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

ker, M.; Wu, D.; Yu, G.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; M01D:21016719

A:Accession: B96716

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 836 <STO>

A:Cross-references: GB:AE005173; NID:q7705100; PIDN:AAF67779.1; GSPDB:GN00141

C:Genetics:

A:Gene: F23010.20

A:Map position: 1

Query Match 30.2%; Score 774.5; DB 2; Length 836;

Best Local Similarity 38.1%; Pred. No. 3.7e-23;

Matches 177; Conservative 73; Mismatches 143; Indels 71; Gaps 12;

QY 4 PPAPSKL-----KLTSE-----DSLTKQ-PEEFVDLEKIGEGSYGSVERKAH 46

Db 209 PSSSSSKLHGDENRKMQQNSMTSLPSDITREDPTTKTEFELNKGSGSYGVARD 268

QY 47 KESGOVAIKQV--ESDLEIKETISIMQCCDSPPVVKYSGYRKNDLIMVWEYCG 103

Db 269 LKTSLEIVAKVSLTEGEGYEIEGELMLOQCHHPVAVYLSYQGEEDLVIMVEYCG 328

QY 104 AGSVSDILRNKLTIEDIATILKSTLGLLEYLHFMKIRHDKAGNINILNTEGHAKLA 163

Db 329 GGSVADLMNTEALFEYGLAVICREALKGLAYLSHYKVRHIDGNGNILLTEGEGYKLG 388

QY 164 DFGVAGQLDFTMAKRNVTIGTPPMAPVYIOELGYNCAVDISLIGITISIMAEKRPYAD 223

Db 389 DFGVAGQLDFTMAKRNVTIGTPPMAPVYIOELGYNCAVDISLIGITISIMAEKRPYAD 448

QY 224 IHPRATFMTPTNPPTFRPELMSDFTDVEYKCLVKNPEGRATQLOHPFIKNAKP 283

Db 449 VHPKRVLFMTSIEPAPLEKESKSLVFHFDFVAKCLTKTEPRLEPTAAEMLKRFVERCKT 508

QY 284 -VSLRLDITAMEITAKRDEOORELEEEENSDEDELDSHTMVKTSVGECC----- 335

Db 509 GASMSKIKTESRQIRATMALQASVAPSL-----DTSLGPKSSSELGITVPSKP 561

QY 336 -----TMRATSTMEGAQTMIEH--NSTMLE-----SDLGTVINSEDEEEDGTMRNATS 385

Db 562 PNSTEADPSTLNR-----QHTGNTVLAGSGDGTIVIGEDFETESRSQVRE 615

QY 386 PVOVRPSPMDYFDKODKRNKSHENCONMHEPPPMASNVYPDW 429

Db 616 KESSSQE-----EGVPRFPGEE--LPDSN 639

# RESULT 4

serine/threonine protein kinase - mouse

C:Species: Mus musculus (house mouse)

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T14157

R:Pykowski, B.; Hicklin, D.J.; Kornhaber, G.; Dellaratta, D.V.; Witte, L.

submitted to the EMBL Data Library, December 1997

A:Reference number: T14157

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1233 <PVT>

A:Cross-references: EMBL:AF039574; NID:q2773155; PID:q2773156; PIDN:AB96682.1

Query Match 27.9%; Score 714.5; DB 2; Length 1233;

Best Local Similarity 32.0%; Pred. No. 1e-20;

Matches 171; Conservative 110; Mismatches 165; Indels 91; Gaps 14;

QY 10 KLKRLSDSLTKQ-----PEEFVDLEKIGEGSYGSVERKAHKEGGOVAIKQV 60

Db 8 KIFELGEEKKKQYEHVYKRDLPNEEWEITIGELGDAPFKYKAQNKERNVLAARVID 67

QY 61 -ESDLEIKETISIMQCCDSPPVVKYSGYRKNDLIMVWEYCGAGSVSDILRNKTL 118

Db 68 KSEELDEYWEIDLASCHPNIVKLDFAFYENNIMLILIECAGADAVALELERL 127

QY 119 IDEIATILKSTLGLLEYLHFMKIRHDKAGNINILNTEGHAKLADFGVAGQLDFTMAKR 178